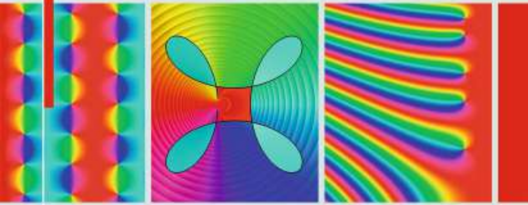


Elias Wegert

Visual Complex Functions

An Introduction with Phase Portraits

Birkhäuser

Birkhäuser

Elias Wegert

Visual Complex Functions

An Introduction with Phase Portraits

Birkhäuser

Elias Wegert
Institute of Applied Analysis
TU Bergakademie Freiberg
Freiberg
Germany

ISBN 978-3-0348-0179-9 ISBN 978-3-0348-0180-5 (eBook)
DOI 10.1007/978-3-0348-0180-5
Springer Basel Heidelberg New York Dordrecht London

Library of Congress Control Number: 2012946174

Mathematics Subject Classification (2010): Primary: 30-01; Secondary: 00A66, 30-02, 30B10, 30B40, 30C10, 30D20, 30D30, 30J10, 33-01, 33-02, 97I20, 97I80

© Springer Basel 2012
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer Basel AG is part of Springer Science+Business Media (www.birkhauser-science.com)

Dedicated to the memory of Lothar von Wolfersdorf

Contents

Preface **ix**

1 Getting Acquainted **1**

2 Complex Functions **13**
2.1 Complex Numbers 14
2.2 Functions and Mappings 23
2.3 Arithmetic and Geometry 25
2.4 The Analytic Landscape 27
2.5 Color Representations 29
2.6 Convergence and Continuity 42
2.7 Some Plane Geometry 45

3 Analytic Functions **59**
3.1 Polynomials and Rational Functions 60
3.2 Power Series 72
3.3 Introduction to Analytic Functions 94
3.4 Analytic Functions in Planar Domains 99
3.5 Analytic Functions on the Sphere 112
3.6 Analytic Continuation 117

4 Complex Calculus **133**
4.1 Complex Differentiation 134
4.2 Complex Integration 150
4.3 Cauchy Integral Formula 169
4.4 Laurent Series and Singularities 175
4.5 Residues 183
4.6 Conjugate Harmonic Functions 191

5 Construction Principles **203**
5.1 Function Sequences 203
5.2 Normal Families 206

5.3 Function Series 210
5.4 Infinite Products 219
5.5 Cauchy Integrals 227
5.6 Integrals with Parameters 241

6 Conformal Mappings **253**
6.1 Mappings of Planar Domains 254
6.2 Special Conformal Mappings 259
6.3 Möbius Transformations 271
6.4 The Riemann Mapping Theorem 278
6.5 Boundary Correspondence 283
6.6 The Reflection Principle 289
6.7 Elliptic Integrals 295
6.8 The Schwarz-Christoffel Formula 302

7 Riemann Surfaces **311**
7.1 Global Analytic Functions 312
7.2 Lifting Techniques 314
7.3 Typical Examples 317
7.4 Analytic Functions and Branch Points 322
7.5 Abstract Riemann Surfaces 330
7.6 Practical Excursion 337

Epilogue **345**

Bibliography **347**

Index **352**

Preface

Are you ready for an experiment? Then please fix the first thought that comes to your mind when you are asked “What is a sine function?” Once you have done that compare your answer with the four options on page xiii.

Did you chose (b)? Indeed most of us memorize pictures more easily than formulas, and an overwhelming majority of people associate with the sine function the wavy curve of its graph.

However, this is a book on complex functions. So let me ask again and more precisely, “What is a *complex sine function?*” Probably this time your inner eye refuses to show you a picture, and this happens for a simple reason – our brain is trained to visualize objects in three spatial dimensions, while the graphs of complex functions live in a four-dimensional space. Hence most of us are unable to imagine such an object.

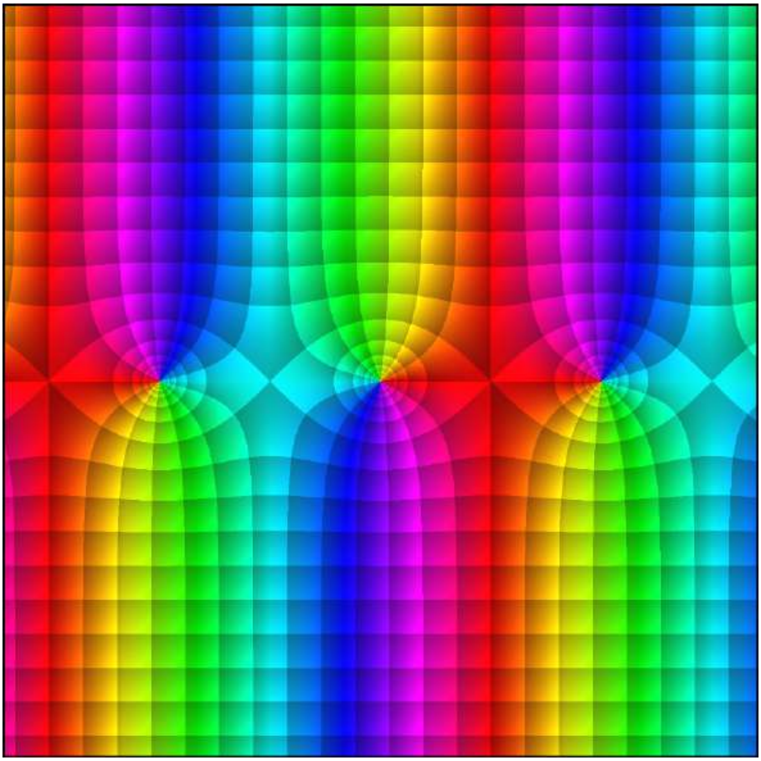


Figure 1: An (enhanced) phase portrait of the complex sine function

But the situation is not completely hopeless since the missing spatial dimension can be added using *color*. The figure above shows an image of the complex sine function, and though you probably cannot interpret this cryptic picture yet, this will change when you have read this book.

What it is about. This textbook is an introduction to theory and applications of complex functions. The presentation is intuitive and requires only basic knowledge of calculus. Volume 1 covers the standard topics of a first course in complex analysis. With a few exceptions all results are provided with proofs. The forthcoming Volume 2 will be devoted to selected topics and various applications of complex methods, like integral transforms, boundary value problems, and signal analysis.

What distinguishes this book from other texts in the first instance is the systematic use of *phase portraits*, a special coloring technique which visualizes functions as images. Readers will learn not only how properties of a function are reflected in and can be read off from its phase portrait, but also how experiments with phase portraits can be designed and adapted to answer specific questions.

Secondly, I have gone about building the theory of complex functions in a manner that is somewhat out of fashion these days. It is well known that the three prominent protagonists of complex function theory in the 19th century, Augustin Louis Cauchy, Bernhard Riemann and Karl Weierstrass, considered the subject from differing points of view. Most contemporary textbooks[1] develop the theory of analytic functions along the ideas of Cauchy and Riemann, with complex differentiability as the entry point, followed by complex integrals. Here we adopt Karl Weierstrass' *constructive* approach via power series, which also best exemplifies the *intrinsic evolution* of mathematical concepts: once the complex number system is established, the interplay between asking questions and looking for answers guides us from one step to the next till we *almost inevitably* end up with Riemann surfaces.

Why to read it. Complex functions are everywhere. Besides their relevance in most fields of mathematics, they have also become indispensable tools in the natural sciences and engineering. It was Carl Friedrich Gauss, an outstanding pioneer of complex analysis, who already wrote in the first half the 19th century:[2]

> *"Die vollständige Erkenntniss der Natur einer analytischen Function muss auch die* Einsicht *in ihr Verhalten bei den imaginären Werthen des Arguments in sich schließen und oft ist sogar letztere unentbehrlich zu einer richtigen Beurtheilung der Gebarung der Function im Gebiete der reellen Argumente."*

> *"Complete knowledge of the nature of an analytic function must also include* insight *into its behavior for imaginary values of the arguments, and often the latter is indispensable for a proper appreciation of the behavior of the function for real arguments."*

Complex functions sometimes have the reputation of being mysterious entities; *seeing* these alien objects may help to overcome the awe one might feel while dealing with them.

---

[1] One of the rare exceptions is Noguchi [46].

[2] Gauss Werke Vol. 10, No 1, p. 405, quoted after [56]

Phase portraits provide functions with an individual face and deepen our intuitive understanding of basic and advanced concepts in complex analysis. They reveal intrinsic structures behind the formulas, literally open our eyes to the wonderful realm of complex functions, and may serve students, teachers, scientists, and engineers as simple and efficient tools in their work.

It is my conviction that even experienced mathematicians can benefit from this visual approach as it might help them to get a fresh perspective on old problems and inspire them to ask new and more challenging questions.[3]

How to read it. We assume that readers have some knowledge of real numbers, calculus, and two-dimensional Euclidean geometry. Though the text is self-contained, it may also serve as a companion to one of the many excellent textbooks on complex analysis; this is certainly an option for those who are interested in phase portraits but do not like Weierstrass' approach. Readers who are already familiar with complex functions may not need to do more than simply browse through the book, look at the illustrations, and pick out themes they find interesting to investigate further.

A number of worked out examples demonstrating the use of phase portraits are scattered throughout the text. More importantly, readers are invited to participate actively and create their own experiments. With a computer and some basic software it is quite easy to generate nice and interesting pictures. For example, making a phase portrait of an elementary function requires less than ten lines of MATLAB® code.[4]

What it contains. After some general comments on the visualization of complex functions and an informal introduction to phase portraits in Chapter 1, the systematic exposition begins in Chapter 2. Having established the system of complex numbers and their arithmetic operations, we investigate general properties of functions and discuss various options for their pictorial representation.

Equipped with the toolkit of phase portraits, we explore and discover analytic functions in Chapter 3, following some 'natural' line of development. We begin with those functions that can be formed using only the four basic arithmetic operations: polynomials and rational functions. In the next step, limit processes lead to power series, which in turn are the local building blocks of general analytic functions. Weierstrass' disk chain method of analytic continuation facilitates the transition from local to global entities. The functions resulting from this procedure however, may be 'multiple-valued', which does not fit into the usual concept of a function. A satisfactory solution, functions on Riemann surfaces, is technically more demanding and will therefore be postponed till Chapter 7. Those unhappy with this proposal may continue with Chapter 7 directly after Chapter 3.

In Chapter 4 we bring in the concept of complex differentiability. It allows us to push the theory much further and helps us to avoid cumbersome manipulations with power series. Here we develop the powerful machinery of complex calculus, including series expansions, path (contour) integrals, integral formulas, and residues.

---

[3] Why do the orbits of the mathematical pendulum show up in the phase portrait of the sine function depicted in Figure 1?

[4] MATLAB is a registered trademark of the MathWorks Inc. Most images in this book have been created using MATLAB.

In Chapter 5 we have collected various techniques for constructing analytic functions: sequences, series, products and integrals. A central theme is the notion of normal convergence, which in connection with Montel's theorem provides us with an efficient tool for verifying the existence of solutions to extremal problems.

The focus of Chapter 6 is on geometric aspects. Here we adopt Riemann's view of complex functions as mappings between (domains of) two complex planes. In this geometric setting, analytic functions are characterized as angle preserving mappings. A large part of the chapter is devoted to bijective conformal mappings between specific domains, in particular to Möbius transformations, elliptic integrals, and Schwarz–Christoffel mappings. Highlights are the Riemann mapping theorem and the Carathéodory–Osgood theorem on boundary correspondence.

Chapter 7 is devoted to Riemann surfaces, and we will look at them at varying levels of abstraction. Concrete Riemann surfaces are constructed from a patchwork of function elements, abstract Riemann surfaces are modelled as manifolds on a topological space. Since phase portraits are images, functions on Riemann surfaces can be depicted directly on the familiar models of such surfaces embedded in three-dimensional space. In the alternative representation of a function on the (flat) sheets of its Riemann surface, phase portraits help us to locate the branch cuts and to explore in which way the different sheets are glued along their edges to form the global surface.

Literature. It is impossible to list the vast existing literature of introductory texts in complex analysis. Within the last ten years alone, more than 25 textbooks in English[5] have been published in this field. In writing this text I benefitted especially from Ablowitz and Fokas [1], Donaldson [9], Freitag [20], Freitag and Busam [21], Henrici [26], Krantz [31],[33], Lin [35], Lorenz [36], Marsden and Hoffman [41], Needham [44], Norton [47], Palka [52], Remmert [56], Schabat [62], and Ullrich [67]. Since many results, proofs and examples are standard, references are not often explicitly mentioned. The list of references at the end is based on the literature which was available and useful to me, and which I can recommend to students. However, this list is by no means complete and could be supplemented by a great number of other excellent texts. An interested reader will certainly have no problems in finding appropriate material for further reading which suits his or her taste better.

A plea for phase. Since this is a book involving phase portraits, let us spend a few moments exploring the role of phase in general. Basically there are two representations of complex numbers, the cartesian form $z = x + \mathrm{i}\, y$ and the polar form $z = r\,(\cos\varphi + \mathrm{i}\,\sin\varphi)$. While the first one uses the real part x and the imaginary part y, the polar form employs the modulus r and the argument φ of z.

---

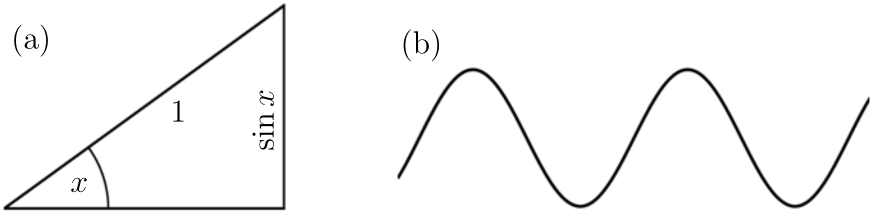
[5] According to Zentralblatt Mathematik.

Among these quantities x, y and r are uniquely determined, but this is not so for φ. This multi-valuedness makes the argument somewhat unwieldy, and students often try to avoid using it.

A simple trick may help to resolve this complication: substitute the argument φ by the phase $\cos\varphi + \mathrm{i}\,\sin\varphi$ whenever possible. Since the phase of z is just $z/|z|$, it is well defined for all complex numbers except zero.

Compared to the modulus, the importance of phase is often underestimated. But the truth is, due to a subtle asymmetry between modulus and phase, the phase of a function sometimes delivers even *more information* than the modulus.[6] Since phase portraits depict the color-coded values of the phase on the domain of the function, it is my hope that their use may contribute to give phase the attention it deserves.

My acquaintance with complex functions dates back almost forty years, but it took a long time until I could begin to *see* my friends. I love them even more ever since I know their phases. This book has been written to let you share my joy.



(c) $$\frac{1}{2\mathrm{i}}\left(\mathrm{e}^{\mathrm{i}x} - \mathrm{e}^{-\mathrm{i}x}\right)$$

(d) $$\sum_{k=0}^{\infty}(-1)^k\frac{x^{2k+1}}{(2k+1)\,!}$$

Figure 2: What do you associate with the sine function ?

Acknowledgements. My first thanks go to Richard Varga. His enthusiastic comments on the usefulness of phase portraits in teaching complex analysis convinced me to undertake this project, the full dimensions of which I could barely foresee at that time.

Several ideas about phase portraits emerged while working on the research project "Nonlinear Riemann-Hilbert Problems, Circle Packing and Conformal Geometry", supported by the Deutsche Forschungsgemeinschaft (DFG) by grant We1704-8/2.

---

[6] This will be made precise in Volume 2, see also Wegert and Semmler [70].

The book has benefitted immensely from the work of Gunter Semmler. He read carefully several stages of the complete text, checked many details, and made proposals for improving the presentation. Even though I could not follow all his recommendations, his valuable suggestions and constructive critical remarks are very much appreciated. The blame for the mistakes that remain is mine.

Daniela Kraus and Oliver Roth read an earlier version of the manuscript and corrected a number of errors. Our conversations supported and encouraged me in the process of writing, and I appreciate their useful comments very much.

Special thanks go to my old friend and colleague Albrecht Böttcher, who made several suggestions for improvements, and helped to overcome a critical period when I got really tired. Moral support came also from Nick Trefethen, who expressed his faith in the value of writing a textbook involving phase portraits and stimulated the project in various ways.

The book would not have been finished in due time without the help of Marko Lindner who took over my teaching responsibilities at the TU Bergakademie Freiberg in the summer term 2011, thereby enabling me to concentrate entirely on the writing of the manuscript.

Our students David Hahn and Daniel Lorenz read through large parts of the manuscript, eliminated errors and proposed changes from a student's perspective. My former colleague Günter Schaar discovered several remaining typos. I also would like to thank Jens Wirth, who ventured to use parts of the book in his introductory lecture on complex analysis.

I am greatly indebted to Meenaxi Barkataki-Ruscheweyh who never hesitated to spend her time with the manuscript. With her professional and dedicated work as copy editor she improved many passages and significantly increased the readability of the text.

Thomas Hempfling, the Managing Director of Springer Basel, immediately supported the project when I approached him with the first idea of the book and he gladly agreed to its transmutation from a small booklet of estimated 150 pages into a two-volume text expected to run into more than 600 pages. I am very grateful that he accompanied all stages of the production and accepted my repeated shifts of "deadlines". It is also a pleasure to thank the staff of Springer/Birkhäuser, in particular Reinhold Joest and Sylvia Lotrovsky, for a pleasant and smooth collaboration. Ines Bruhn's professional assistance in preparing some graphic files for publication was also most valuable.

Last but not the least, I would like to express my gratitude to my wife Petra for her infinite patience and for allowing me to ignore almost all duties of daily life while I was writing, as well as to all friends who showed interest in the progress of the ongoing project and gave me emotional support.

Freiberg, June 2012 Elias Wegert

A picture is worth a thousand words.

Folklore

Chapter 1

Getting Acquainted

Graphical representations of functions belong to the most useful tools in mathematics and its applications. While graphs of (scalar) real-valued functions can be depicted easily in a plane, the graph of a complex function in one variable is a surface in four-dimensional space. Since our imagination is trained in three dimensions, most of us have difficulties in 'seeing' such an object.

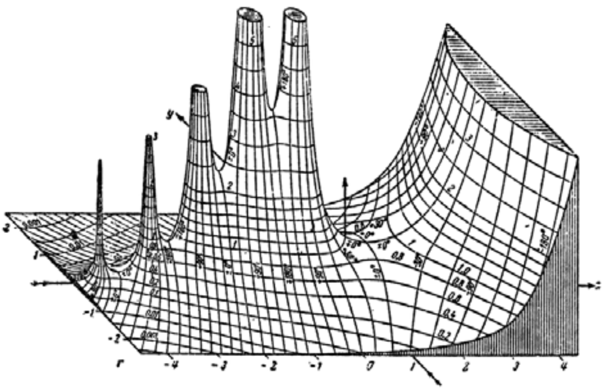


Figure 1.1: A historical analytic landscape

A traditional concept for visualizing complex functions is the so-called *analytic landscape*. Probably introduced by Edmond Maillet [40] in 1903, it depicts the graph of the absolute value of a function. In the first half of the preceding century analytic landscapes became rather popular. Figure 1.1 reproduces a historical illustration from the book [27] by Jahnke and Emde of 1909. It shows the analytic landscape of the complex Gamma function and reached an almost iconic status. Today it is hard to believe that this detailed hand-drawn picture could be created without the help of computers!

Analytic landscapes involve only the *modulus part* of a function, the *argument part* is lost. In the era of black-and-white illustrations this shortcoming was often compensated by endowing the analytic landscape with lines of constant argument

E. Wegert, *Visual Complex Functions: An Introduction with Phase Portraits*,
DOI 10.1007/978-3-0348-0180-5_1, © Springer Basel 2012

as in Figure 1.1, where the argument is explicitly indicated by its numerical value. Modern computer technology allows us to achieve the same effect much better using *colors*, which yields the *colored analytic landscape* shown in Figure 1.2. One should not worry about the correct interpretation of colors here, this will be explained in some detail in Section 2.5, but comparing the figures reveals that the isochromatic lines of Figure 1.2 correspond to the lines of constant argument in the historical Figure 1.1.

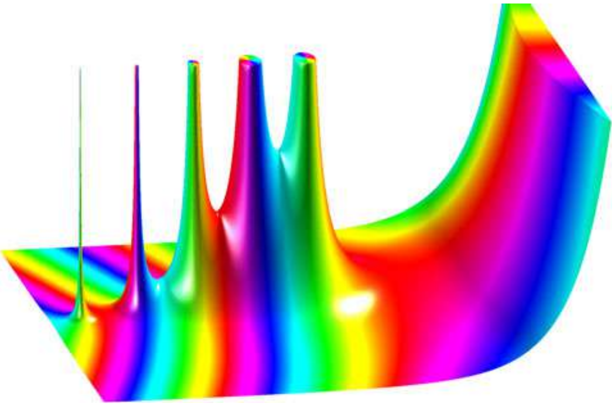


Figure 1.2: Colored analytic landscape of Γ

With colored analytic landscapes the problem of depicting complex functions could be considered solved. But if there *is* such a nice way of visualization, then why are many textbooks on complex analysis without a single picture of a function?

The first reason why even professional mathematicians have never seen the entities they are working with is that *beautiful* analytic landscapes are not easy to generate, even with contemporary software.[1] And, honestly, does one learn so much more from simply seeing the towers?

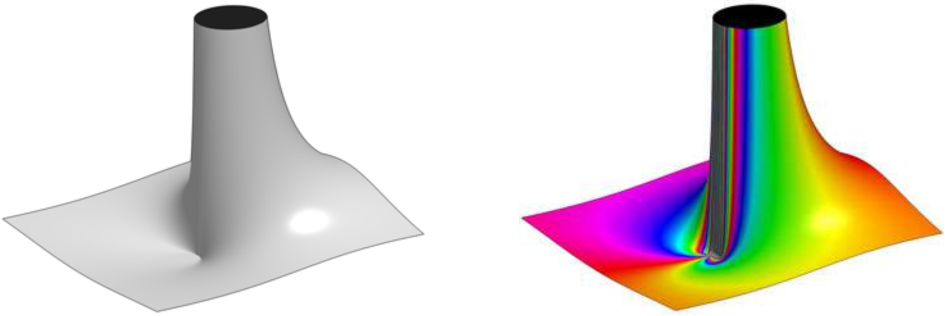


Figure 1.3: Plain and colored analytic landscape of $f(z) = \exp(1/z)$

In fact there are situations where the use of plain analytic landscapes, depicting only the modulus of a function, can be misleading. For example, the analytic landscape of the function $f(z) = \mathrm{e}^{1/z}$ shown in Figure 1.3 (left) does not differ very much from one with a pole and a neighboring zero, though the behavior of the function is quite different and much more involved, as can be seen in Figure 1.3 (right) which has the phase information added.

---

[1] Rich sources of analytic landscapes are the Digital Library of Mathematical Functions [49], the accompanying Handbook of Mathematical Functions [50], the Atlas of Functions [48], and the Wolfram Functions Site [72].

Finally, everybody who tried to use analytic landscapes seriously knows that it might be rather difficult to read off the information one is interested in. The range of the modulus of a function can be quite large, and often essential parts are hidden in the valleys behind mountain rims or covered by the towers of poles.

The best one can do in such situations is to look at the colored analytic landscape straight from the top. The result is a flat color image, as is shown in Figure 1.4 for the Gamma function. This *phase portrait* represents the color coded phase (or argument) of the function. Like the plain analytic landscape, the phase portrait depicts only 'one half' of a function, but it turns out that the phase is better suited than the modulus to understand a function and to reconstruct its properties. And one goal of this book is to convince you that it is really so.

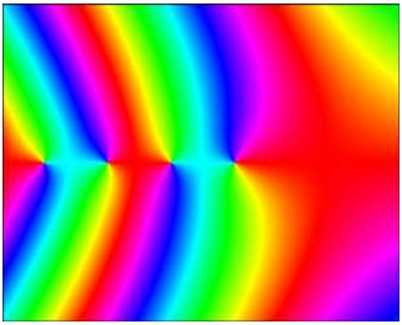


Figure 1.4: A phase portrait of Γ

Even if one does not know how phase portraits are constructed, it is clear that the functions depicted on this page are quite different, and with some intuition one can perhaps even guess some of their properties.

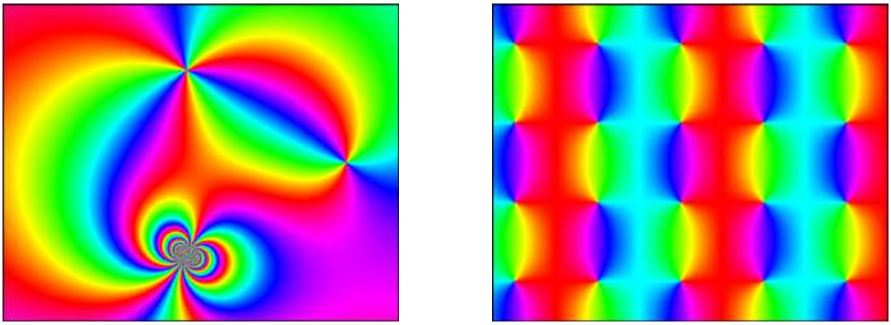


Figure 1.5: What can we learn from these phase portraits?

How phase portraits can be decoded is a theme which runs parallel to the main course of the book. Since the reader might be curious to see how specific properties of a function are reflected in its phase portrait, we give some informal introduction right now.

A *plain phase portrait* of a complex function f is an image which depicts the color-coded phase $f/|f|$ on the domain of f. Since the values of the phase lie on the complex unit circle, they can be represented on a ‘color wheel’. The phase portraits in Figures 1.4 and 1.5 are constructed in this manner, using the standard `hsv` color scheme. Of the three parameters *hue*, *saturation*, and *value*, only the first one is involved, representing saturated colors on a circle, as in the Figure 1.6 (left). The picture in the middle shows the color-coded phase of points in the complex plane.

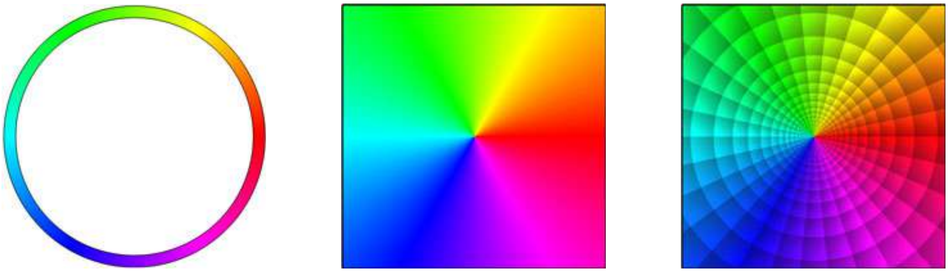


Figure 1.6: The color wheel, a plain and an enhanced phase portrait of $f(z) = z$

Though phase portraits do not take into account the *modulus* of a function, they contain enough information to reconstruct *analytic* functions uniquely up to a positive factor. Nevertheless, we shall often resort to *enhanced* phase portraits, which allow one to read off particular features more easily.

Figure 1.6 (right) depicts such an image of $f(z) = z$, where pure colors (decoding phase) are overlayed by shades of gray. An enhanced phase portrait of a more complicated function is shown in Figure 1.7. How these pictures are generated is discussed at length in Section 2.5, here we only give some intuitive description.

First of all we remark that the discontinuities of the gray-shading form two families of lines. One family consists of *isochromatic lines*, along which f has *constant phase*. The second family, perpendicular to the first one, is formed by (logarithmically spaced) *contour lines* of the absolute value $|f|$. Both families together subdivide the square depicted into ‘tiles’ of different shapes and sizes.

Looking at Figure 1.7 we immediately discover a few exceptional points where all colors meet, three such points are marked Z, P and E. Comparing first Z and P, we see that near Z every color appears only once, while the color circle is run through twice when we move around P. Another difference between the two points is the orientation of colors. Walking around Z in the counter-clockwise direction, we meet the colors in the order red, yellow, green, blue, while this order is reversed at P. The explanation is that Z is a *simple zero* of f, while P is a *double pole*.

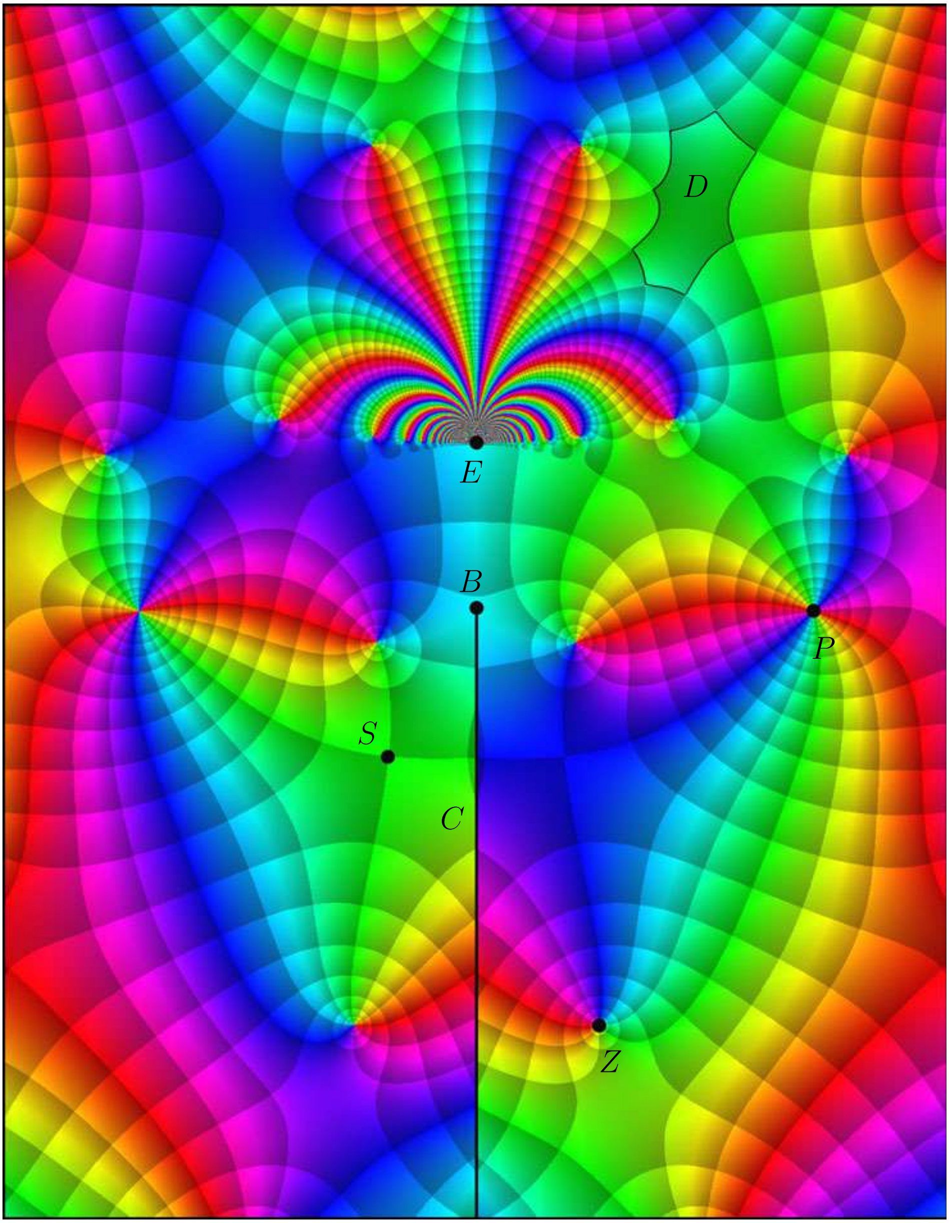


Figure 1.7: The enhanced phase portrait of an analytic function

The behavior of the function near E is obviously rather wild. This point is an *essential singularity*. We shall see in Section 4.4 that in any neighborhood of an essential singularity the function attains almost all complex numbers as values.

Another interesting observation is the crossing of two (green) isochromatic lines at point S. Such points are called *saddle points*, they appear at the zeros of the derivative f' of f. From the structure of the isochromatic lines through S one can read off that S is a simple zero of f'. There are other saddle points which are hidden in some amoeba-like tiles, for example in the region marked D. Though the isochromatic lines through this point are not enhanced, and thus the precise location of the saddle cannot be seen, the order of the zero of f' can be determined from the shape of the tile D in which it is contained. Can you guess it?

Along the black line C (which has been inserted for clear visualization) the colors on the left side do not fit with the colors to the right. This line of discontinuity is a *branch cut*. If we would fix its endpoint B, which is a *branch point* of the function f, and move the cut to the right, we would see that the left part of the phase portrait in fact extends to the right, and vice versa. The function f admits an *analytic continuation* across the cut. This observation gives the clue to the construction of *Riemann surfaces*.

The phase portrait also contains information on the *growth* of the function. The *isochromatic lines* coincide with the line of *steepest descent* (or ascent) of $|f|$. Walking along such a line, we cross the (perpendicular) family of *contour lines*, i.e., the lines of constant modulus $|f|$. Between two consecutive crossings the modulus of f is always *changing by the same factor* c. The color scheme is chosen such that the value of c can be read off from the phase portrait; it is $c = \mathrm{e}^{2\pi/n}$, where n denotes the number of enhanced isochromatic lines of different color. In the example at hand there are nine such lines (which can be conveniently counted at the point Z, for instance), so that $c = 2.009993\ldots$ is almost 2. In which direction $|f|$ decreases or increases can be seen from the *ordering* of the colors. For example, travelling along a yellow line with red to the right and green to the left, one moves in the direction of increasing $|f|$. So we see that $|f|$ increases towards the upper right corner of the rectangle depicted. While the growth of $|f|$ is still moderate in that region, it is much faster when we approach the point E from its 'upper' ('wild') side. In contrast to this, the function behaves quite tamely in the region 'below' E.

The reader may wonder why the growth of f was related to the number of different isochromatic lines and not to the contour lines of $|f|$. The reason is that the parameters of the enhanced phase portraits are chosen such that the 'densities' of both families coincide. But while one usually can count easily how many enhanced isochromatic lines of different colors a phase portrait has (for instance in a neighborhood of a zero or a pole), similar information for the contour lines is not accessible directly.

That phase portraits with equal densities of isochromatic lines and contour lines can be at all constructed rests on a special property of analytic functions, which also manifests itself in the *shape of the tiles*. Inspection of the phase portrait in Figure 1.7 reveals that almost all tiles are curvilinear quadrilaterals with right-angled corners and almost equal side lengths. A detailed explanation of this effect will be given in Chapter 6, it suffices here to say that it is based on the *conformality* of the mapping $z \mapsto w := f(z)$, which means that the oriented intersection angles between curves in the z-plane and their image curves in the w-plane are always the same.

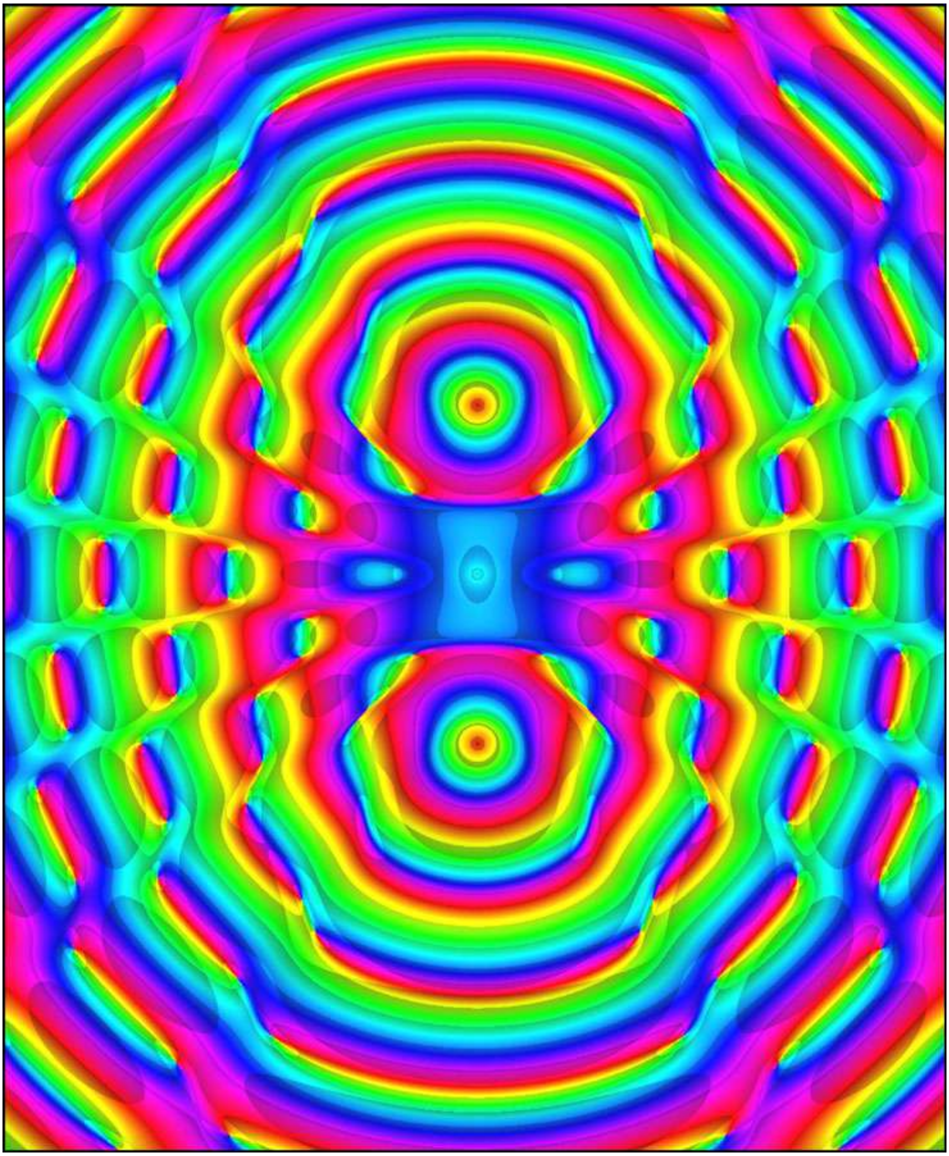


Figure 1.8: The enhanced phase portrait of a non-analytic function

Figure 1.8 shows an enhanced phase portrait of a function which is not analytic. It visualizes the superposition of three (complex) circular waves emerging from three points on the vertical symmetry axis of the rectangle. The ‘upper’ and the ‘lower’ waves are synchronized, while the ‘middle’ one has a lower frequency. The general appearance of this phase portrait is quite different from Figure 1.7, for instance it lacks the typical rectangular tiling pattern, in some regions the isochromatic lines are almost parallel to the contour lines.

Practical applications of phase portraits will be presented in Volume 2. In order to give a idea of what it is about, we sketch an example from signal processing.

An *analog signal* is a (real or complex valued) function $f = f(t)$ on the real line (in an appropriate function space). The function f is also referred to as the *time-domain representation* of the signal. Signals can be decomposed into harmonic oscillations $\mathrm{e}^{\mathrm{i}\omega t}$ by means of the Fourier (or Laplace) transform, which associates with each frequency ω the magnitude $a(\omega)$ and the phase $\varphi(\omega)$ of the signal. The result of this spectral analysis is a *frequency-domain representation*

$$F(\omega) = a(\omega)\,\mathrm{e}^{\mathrm{i}\varphi(\omega)}$$

of the signal, which is said to be the *spectrum* of f. Under appropriate assumptions the spectrum F of a signal extends from the real line to an analytic function in some domain, for instance in the upper half-plane.

Signals can be processed in various ways. The operators which do such transformations can often conveniently be described in the frequency domain. For example, a *filter* is a system which operates such that the spectrum F of the input signal and the spectrum G of the output signal G are related by a simple multiplication $G(\omega) = H(\omega)\,F(\omega)$. The function H characterizes the performance of the filter and is said to be its *transfer function.* More precisely, the modulus of $|H(\omega)|$ is the *amplification* ($|H(\omega)|^2$ is the *gain*) and the argument $\arg H(\omega)$ is the *phase shift* induced by the filter operating on a harmonic oscillation with frequency ω. Transfer functions of filters are *analytic* in a half-plane (which is the upper half-plane in case of the Fourier transform and the right half-plane for the Laplace transform); typically they are even rational functions.

Filters have to be designed for serving various purposes, and often requirements are contradictory. For example, the transfer function of an *ideal low pass filter* should fulfil $H(\omega) = 1$ for $|\omega| < \omega_0$ (pass-band) and $H(\omega) = 0$ for $|\omega| > \omega_0$ (stop-band), but no transfer function satisfying these conditions can be realized. So filter design is an optimization problem where one has to find a compromise between conflicting requirements.

The first mathematical construction of an approximate low pass filter is due to Stephen Butterworth in 1930. The transfer function of a *Butterworth filter* of order n is

$$H(z) = \prod_{k=1}^{n} \frac{1}{(z/\omega_0) - z_k}, \qquad z_k := \mathrm{i}\,\exp\left(\mathrm{i}\,\frac{2k-1}{2n}\right).$$

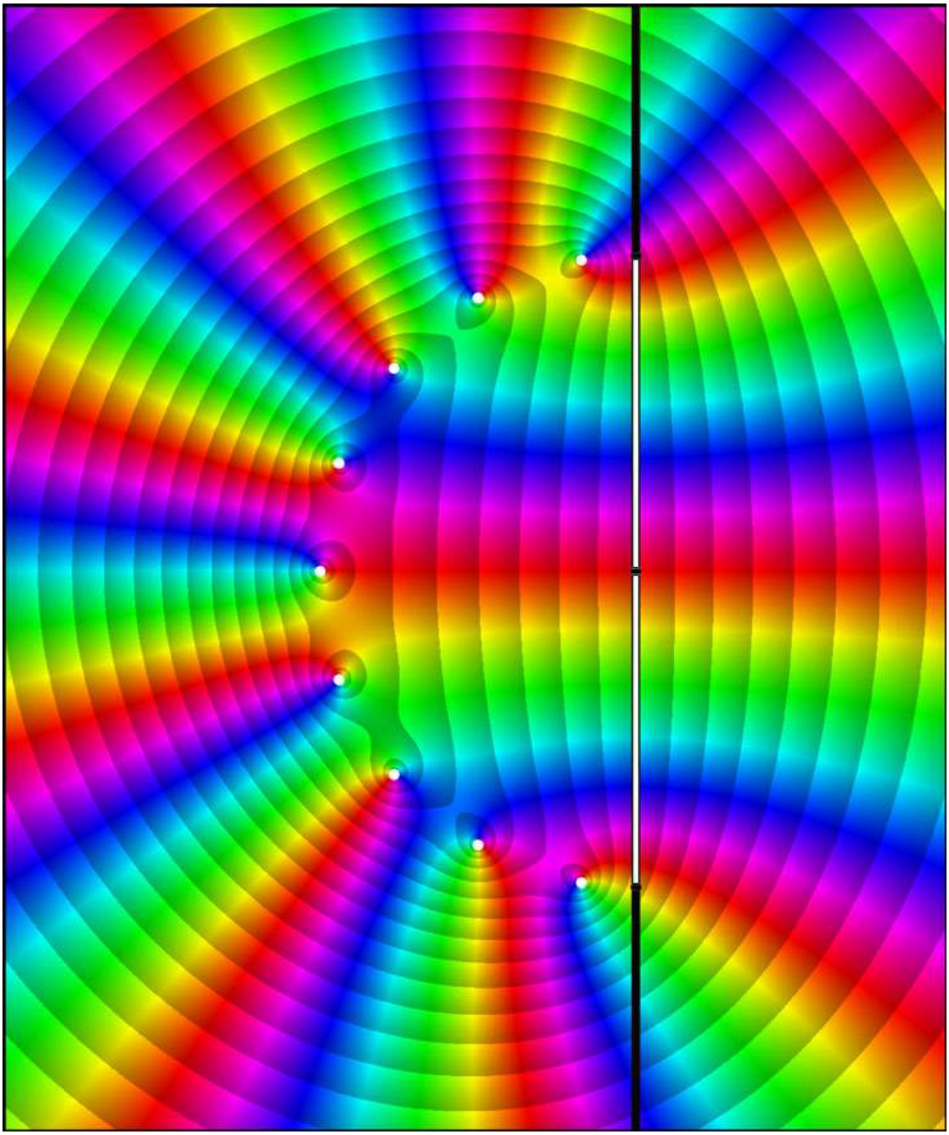
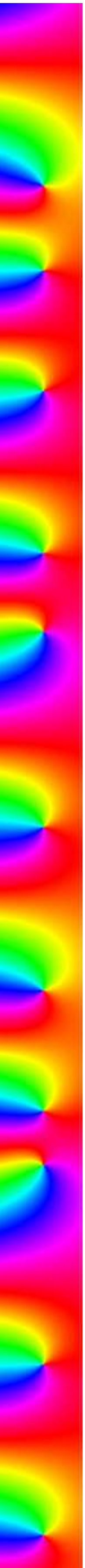


Figure 1.9: Phase portrait of a Butterworth filter with contour lines of $|H|$

Figure 1.9 depicts the phase portrait with enhanced contour lines of $|H|$ of the transfer function of a Butterworth filter of order 9. Since the Laplace transform is more common in applications than the Fourier transform we have adopted here the usual convention to rotate the complex plane such that H is analytic in the right half-plane. Consequently the frequency response is the function $\omega \mapsto H(\mathrm{i}\omega)$.



The nine equally spaced white dots are the poles of H at the points $z_1, \ldots, z_9$, the black dot in their center is the origin. The black-and-white vertical line is the imaginary axis on which the frequency response lives. The white segment corresponds to the pass-band, the black segments represent the stop-band(s) of the ideal low pass filter. In the white zone the amplification $|H|$ is almost constant and equal to one. Along the black segments the modulus decays with increasing $|\omega|$. Between two neighboring contour lines of $|H|$ the modulus of H changes approximately by a factor of 2. The colors represent the phase shift.

Figure 1.10 visualizes amplification (left) and phase shift (right) of the same filter as functions of the frequency. Note that the horizontal and the vertical axes are scaled logarithmically. These graphical representations are called *Bode plots*. After some training in interpreting phase portraits these diagrams can be read off from Figure 1.9.

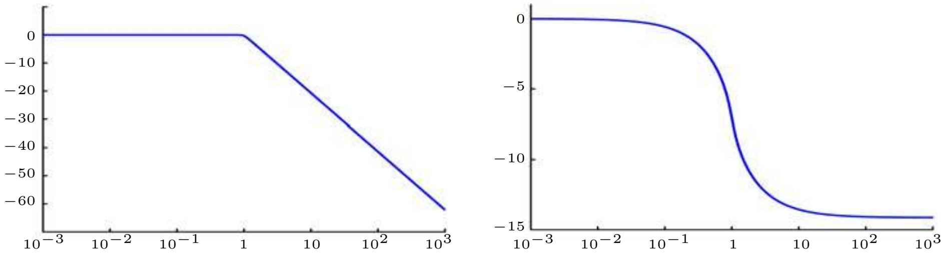


Figure 1.10: Bode plots of Butterworth filter: gain (left) and phase delay (right)

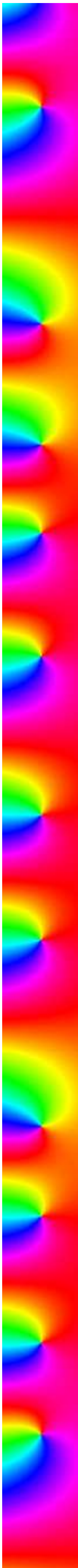
Here our short excursion with phase portraits comes to an end. Before we begin the systematic investigation of complex functions, two general remarks may be helpful.

The first one is a warning: never forget that phase portraits are pictures, and pictures can be misleading. For example, essential details may not be visible because of low resolution, or the values of a function can lie in a range where the phase is almost constant. Using phase portraits for exploring a function requires critical thinking about what one observes.

The second comment is to share a simple trick with you which often helps to make dull phase portraits more interesting.

Adding a constant is usually considered as a minor change to a function, but it may (and usually does) change its phase portrait completely. For example, if the constant c is sufficiently large, then the phase portrait of $f + c$ will be almost monochromatic, which is typically not what we want. On the other hand, adding an *appropriate* constant can be quite useful. For instance, if a phase portrait of a function f is almost monochromatic, choose some point z_0 in the region of interest and consider the function $f - f(z_0)$. Since this function is small near z_0, its phase portrait has a high resolution and acts like a microscope focused at this point.

While reading this book you may wonder about the colored stripes in the margins of some pages (like this one). These are phase portraits of one of the most fascinating functions, the *Riemann Zeta function*, which will be introduced in Section 5.3 and studied further in Section 5.6 and in Volume 2. An explanation of what these images show is given on page 249, here we only remark that they are related to the famous *Riemann hypothesis*, which is one of the seven Millennium Prize Problems. The correct solution to any one of these problems will be honored by a 1,000,000 US-dollar prize to be awarded by the Clay Mathematics Institute.



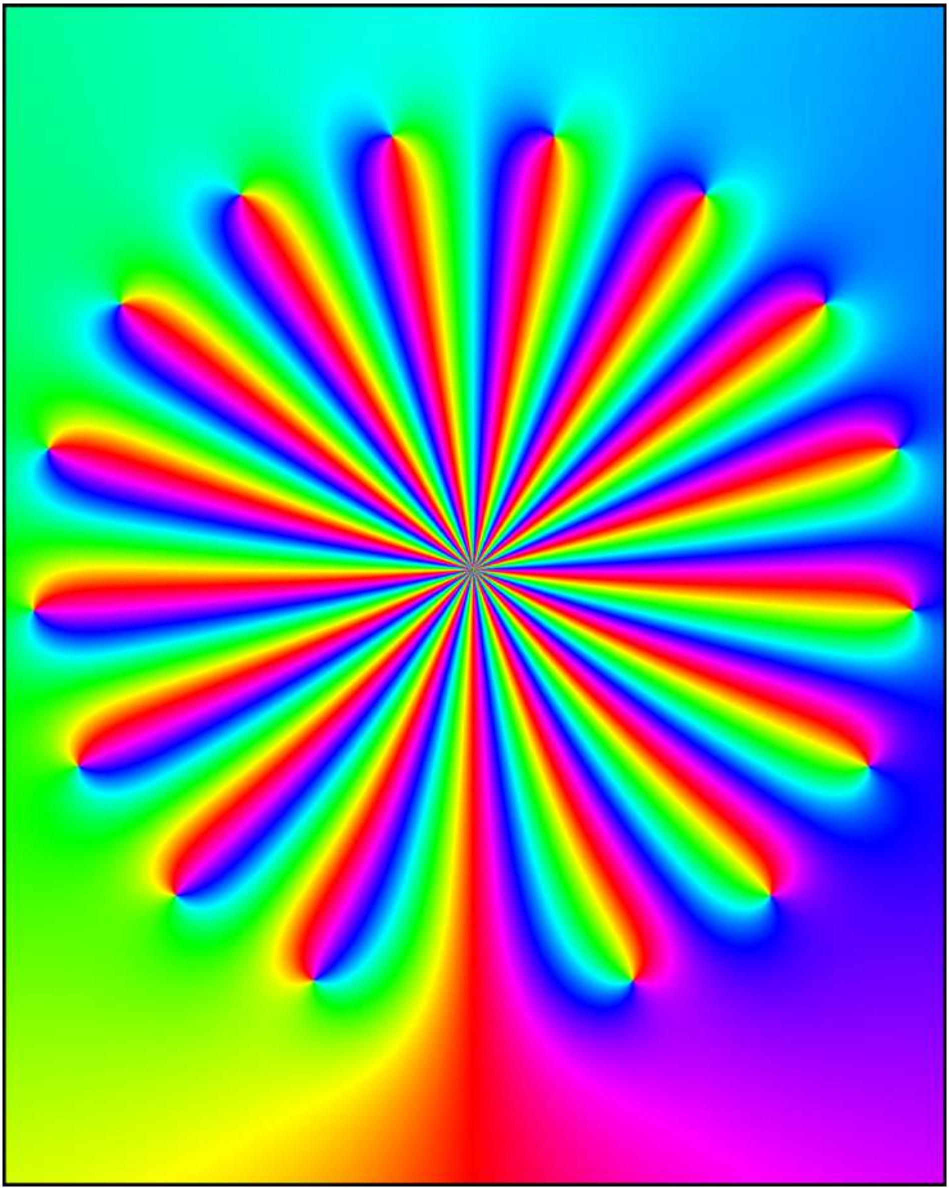


Figure 2.1: Phase portrait of the function $f(z) = \big((\mathrm{i}z)^{17} - 1\big)/\big(\mathrm{i}z - 1\big)$

Chapter 2

Complex Functions

The functions we shall be exploring in this book are complex-valued functions of a single complex variable. When we speak of *complex functions*, we do not necessarily mean that these functions are analytic, although emphasis will be placed on this special and extremely important class.

We begin with a condensed introduction to the arithmetic and geometry of the complex number system. This is not intended as a full course on the subject; we shall give only a brief summary of some fundamental facts and fix some notation. Readers who have never worked with complex numbers before are advised to supplement the study of this text by consulting one of the more comprehensive textbooks, like Krantz [19], Lin [35], Marsden and Hoffmann[41], Needham [44], Palka [52], to mention a few.

The second section reviews some basic material about functions in general, with emphasis on complex functions. This topic is continued in Section 2.3, where the focus moves to geometric aspects of complex functions as mappings between (subsets of) two copies of the complex plane. In particular we investigate the arithmetic operations as geometric transformations.

In Section 2.4 we briefly discuss a classical approach for visualizing complex functions: the analytic landscape or module surface. Although traditionally it was used to depict only the modulus of a function as a graph, the missing information about the argument can be conveniently incorporated by coloring this surface.

In Section 2.5 we lay the groundwork for the central topic of this book: visualization and exploration of complex functions by color representations. We begin by briefly touching upon the method of domain coloring, which depicts a function on its domain by a color representation of its values. A similar idea is utilized in *phase portraits*, which we introduce and discuss next. In contrast to conventional domain coloring, these images represent the color-coded phase (argument) of a function only and neglect its modulus. Nevertheless, the phase portrait contains enough information to reconstruct an *analytic function* uniquely up to a scaling factor.

E. Wegert, *Visual Complex Functions: An Introduction with Phase Portraits*,
DOI 10.1007/978-3-0348-0180-5_2, © Springer Basel 2012

Quite a number of properties of a function can immediately be read off from its phase portrait. Modifications of the underlying color scheme allow us to visualize additional information, which further improves the readability. The section ends with a discussion about the construction of phase portraits on the Riemann sphere by stereographic projection. This part can be skipped on a first reading.

In Section 2.6 we briefly consider sequences of complex numbers and study their convergence in the complex plane and on the Riemann sphere. We also review some elementary properties of continuous functions.

Finally, in Section 2.7, some relevant facts from plane geometry are summarized. Special attention is paid to the concepts of homotopic paths and winding numbers.

We will assume that the reader is familiar with the fundamental facts of arithmetic, linear algebra, real analysis, and plane topology. Readers with some background in complex numbers may skip the first three sections and go directly to Section 2.4.

All notation will be explained where it appears for the first time, here we just mention that $\mathbb{N}$, $\mathbb{Z}_+$, $\mathbb{Z}$, $\mathbb{R}$, and $\mathbb{R}_+$ denote the sets of natural numbers (including zero), positive integers, integers, real numbers, and positive real numbers, respectively.

2.1 Complex Numbers

The Euclidean Plane. Our starting point is $\mathbb{R}^2$, the set of ordered pairs (x, y) of real numbers. Interpreting x and y as Cartesian coordinates of a point in the plane, $\mathbb{R}^2$ is said to be the *Euclidean plane.* We shall therefore identify points in the plane and pairs (x, y) of real numbers.

Arithmetic Operations. Next we introduce *arithmetic operations* with points in $\mathbb{R}^2$. The *sum* and the *product* of (x_1, y_1) and (x_2, y_2) are defined as follows:[1]

$$(x_1, y_1) + (x_2, y_2) := (x_1 + x_2, y_1 + y_2), \tag{2.1}$$

$$(x_1, y_1) \cdot (x_2, y_2) := (x_1x_2 - y_1y_2, x_1y_2 + y_1x_2). \tag{2.2}$$

While the rule for addition is familiar from the corresponding vector operations, the multiplication rule looks a bit odd. It was the Irish mathematician Rowan William Hamilton who discovered that the above definition of multiplication is the only one which is compatible with (vector) addition. Here "compatible" means that both operations together satisfy the ordinary arithmetic rules known from the real numbers, namely the associative, commutative and distributive laws. We shall not verify them here, but readers who are seeing this definition for the first time are encouraged to check them.

---

[1] In (2.1), (2.2), and throughout the text, the sign := indicates a definition: the object on the left-hand side is defined by the expression on the right.

The pairs $(0,0)$ and $(1,0)$ are the *neutral elements* of addition and multiplication, respectively, i.e., for all $(x,y) \in \mathbb{R}^2$

$$(x,y) + (0,0) = (x,y), \qquad (x,y) \cdot (1,0) = (x,y).$$

Further, the pairs of type $(x,0)$ obey the simple arithmetic rules

$$(x,0) + (y,0) = (x+y,0), \qquad (x,0) \cdot (y,0) = (xy,0),$$

which allows one to identify $(x,0)$ with the corresponding real number x, i.e., $x = (x,0)$. In particular the pairs $(0,0)$ and $(1,0)$ represent the real numbers 0 and 1, respectively.

The Imaginary Unit. Besides the two special numbers zero and one, the third distinguished element is the *imaginary unit* $\mathrm{i} := (0,1)$. This extraordinary object satisfies the identity

$$\mathrm{i}^2 := \mathrm{i} \cdot \mathrm{i} = (0,1) \cdot (0,1) = (-1,0) = -1,$$

which demonstrates that the square of a complex number can be negative. The idea of denoting the imaginary unit by the symbol i dates back to a paper by Leonhard Euler in 1733.

Using $x = (x,0)$, $y = (y,0)$, and the definition of multiplication, we obtain $(x,y) = (x,0) + (0,1) \cdot (y,0) = x + \mathrm{i} \cdot y$, which is the *complex representation* of the pair (x,y). Omitting the dot denoting multiplication we obtain the conventional form $x + \mathrm{i}\, y$ of writing *complex numbers*. Usually complex numbers are denoted by a single letter, as in

$$z := x + \mathrm{i}\, y.$$

Because the product is commutative, there is no difference in representing a complex number as $x + \mathrm{i}\, y$ or $x + y\, \mathrm{i}$.

The Complex Plane. Any complex number $z = x + \mathrm{i}\, y$ with $x, y \in \mathbb{R}$ can be identified with the point in the Euclidean plane $\mathbb{R}^2$ having Cartesian coordinates x and y. In this context we refer to $\mathbb{R}^2$ as the *complex plane*, or the *Gaussian plane*, and denote it by $\mathbb{C}$.

The coordinates x and y are said to be the *real part* and the *imaginary part* of z, respectively, and written as $x = \operatorname{Re} z$ and $y = \operatorname{Im} z$. Note that the imaginary part of $x + \mathrm{i}\, y$ is the *real* number y, and *not* $\mathrm{i}y$ as its name might suggest.

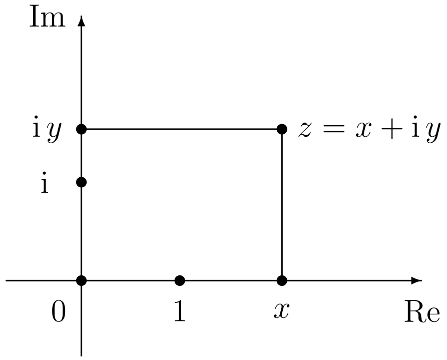


Figure 2.2: The complex plane

We point out that $z = x + \mathrm{i}\, y$ does not necessarily imply that x is the real part and y is the imaginary part of z, this only follows if x and y are real. However, here and also later we shall often automatically assume that x and y are both real without explicit mention.

Complex numbers with $\operatorname{Re} z = 0$ are called *purely imaginary*, the *imaginary axis* is the set of all purely imaginary numbers, the *real line* $\mathbb{R}$ is identified with the *real axis* consisting of all complex numbers with vanishing imaginary part.

The unfortunate attribute "imaginary" enshrouds complex numbers with some kind of mystery. In fact, in the construction of number systems, the transition from real to complex numbers is only a small step which, as we have seen, causes no serious problems once the right approach is found. On the other hand, even after a long time of historical development, the foundation of the real number system still remains rather technical. So the *real numbers* are actually the nontrivial objects which still have their deep secrets. Given real numbers, complex numbers are just pairs of reals, endowed with some additional arithmetic.

More Arithmetic. Indeed, the arithmetic rules are the reason for writing complex numbers in their usual form $x + \mathrm{i}\, y$, and not as pairs (x, y). Remembering that the imaginary unit exhibits the fundamental property $\mathrm{i}^2 = -1$ is all what one needs to compute the sum and the product of complex numbers: just treat i as a variable, apply the arithmetic rules known from the reals, and simplify the result so that it gets the form $x + \mathrm{i}\, y$.

In the next step *subtraction* and *division* are introduced as the *reverse operations* of addition and multiplication, respectively. The *difference* $z_1 - z_2$ of $z_1 = x_1 + \mathrm{i}\, y_1$ and $z_2 = x_2 + \mathrm{i}\, y_2$ is the unique number z which solves the equation $z_2 + z = z_1$, i.e.,

$$z = (x_1 + \mathrm{i}\, y_1) - (x_2 + \mathrm{i}\, y_2) := (x_1 - x_2) + \mathrm{i}\, (y_1 - y_2).$$

Similarly, if $z_2 \neq 0$, the *quotient* z_1/z_2 of z_1 and z_2 is the solution z of $z_2 \cdot z = z_1$. It can easily be verified that this solution is unique, namely

$$z = \frac{x_1 + \mathrm{i}\, y_1}{x_2 + \mathrm{i}\, y_2} := \frac{x_1 x_2 + y_1 y_2}{x_2^2 + y_2^2} + \mathrm{i}\, \frac{y_1 x_2 - x_1 y_2}{x_2^2 + y_2^2}. \tag{2.3}$$

The number $-z := 0 - z$ is termed the *negative* of z, and the *inverse* of $z \neq 0$ is defined by $z^{-1} := 1/z$.

There is a simple trick to remember the division formula (2.3). It utilizes the *conjugate* $\overline{z}$ of a complex number $z = x + \mathrm{i}\, y$, defined by

$$\overline{z} := x - \mathrm{i}\, y.$$

Since $z\,\overline{z} = x^2 + y^2$ is real, multiplying the numerator and denominator of z_1/z_2 by $\overline{z}_2$ allows one to separate the real and imaginary parts of the quotient, which yields (2.3).

The following arithmetic rules for conjugation of complex numbers can easily be verified:

$$\overline{\overline{z}} = z, \quad \overline{z_1 \pm z_2} = \overline{z_1} \pm \overline{z_2}, \quad \overline{z_1\, z_2} = \overline{z_1}\, \overline{z_2}, \quad \overline{z_1/z_2} = \overline{z_1}/\overline{z_2}.$$

We also mention two useful formulas expressing the real and imaginary parts of z in terms of z and $\overline{z}$,

$$\operatorname{Re} z = \frac{z+\overline{z}}{2}, \qquad \operatorname{Im} z = \frac{z-\overline{z}}{2\mathrm{i}}.$$

Integer powers z^n are defined for positive n as the n-fold product of z with itself, and by $z^n := 1/z^{-n}$ if n is negative and $z \neq 0$. Finally, if $z \neq 0$ we set $z^0 := 1$.

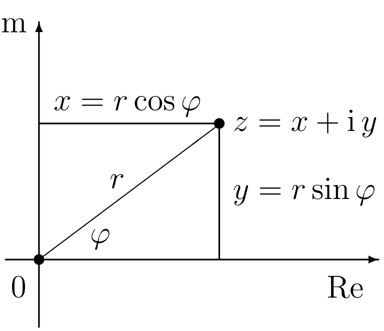


Figure 2.3: Polar representation

Polar Representation. Like points in the plane, complex numbers have an alternative description in terms of *polar coordinates*. If $z = x + \mathrm{i}\, y$ is a complex number different from zero, the real and imaginary parts of z are given by the formulas

$$x = r\, \cos\varphi, \quad y = r\, \sin\varphi, \tag{2.4}$$

where r denotes the distance of the point z from the origin 0, and φ is the oriented angle between the positive real axis and the ray from 0 to z (see Figure 2.3). Inserting (2.4) in $z = x + \mathrm{i}\, y$ yields the *polar representation* of complex numbers

$$z = r\,(\cos\varphi + \mathrm{i}\, \sin\varphi). \tag{2.5}$$

If $z = 0$, then (2.5) holds with $r = 0$ and any value of φ. The value of r is called the *modulus* of z and denoted by $|z|$. The notions *absolute value* or *magnitude* are commonly used as synonyms for modulus. The modulus can be expressed in terms of the real and imaginary part using Pythagoras' theorem,

$$r = |z| = \sqrt{x^2 + y^2} = \sqrt{(\operatorname{Re} z)^2 + (\operatorname{Im} z)^2}.$$

For $z \neq 0$, each real number φ for which (2.5) holds is said to be an *argument* of z. The argument of $z = 0$ is undefined. *Polar angle* or just *angle* are alternative names for the argument of a complex number. Any argument φ of $z = x + \mathrm{i}\, y$ with $x \neq 0$ satisfies

$$\tan\varphi = y/x = \operatorname{Im} z / \operatorname{Re} z. \tag{2.6}$$

However, the *converse is not true* since $\tan\varphi = \tan(\varphi + \pi)$. Consequently the relation (2.6) alone is not sufficient to find an appropriate value of φ, one must take into account the location of z in the four quadrants.

Moreover, even (2.4) does not determine φ uniquely since addition of an integer multiple of 2π does not change the values of the sine and cosine functions.

Argument and Phase. In this text the notation arg z is used to designate an arbitrary argument of z, which means that arg z is a set rather than a number. In particular, the relation arg $z_1 =$ arg z_2 is *not an equation*, but expresses equality of two sets.

As a consequence, two non-zero complex numbers $r_1\big(\cos\varphi_1 + \mathrm{i}\,\sin\varphi_1\big)$ and $r_2\big(\cos\varphi_2 + \mathrm{i}\,\sin\varphi_2\big)$ are equal if and only if

$$r_1 = r_2, \quad \text{and} \quad \varphi_1 - \varphi_2 \in 2\pi\,\mathbb{Z}, \tag{2.7}$$

the latter meaning that there exists an integer k such that $\varphi_1 = \varphi_2 + 2k\,\pi$.

In order to make the argument of z a well-defined number, it is sometimes restricted to the interval $(-\pi, \pi]$. This special choice is called the *principal value* or the *main branch* of the argument and is written as $\operatorname{Arg} z$. Note that there is no general convention about the definition of the principal value, sometimes its values are supposed to be in the interval $[0, 2\pi)$. This ambiguity is a perpetual source of misunderstandings and errors.

Moreover, restricting the argument to its principle value has serious disadvantages and we shall see applications where this is even impossible. One way out is to avoid the argument and to work with the *phase* $\psi(z) := z/|z|$ of z. We shall discuss this issue in more detail later.

Arithmetic Revisited. Let us now briefly reconsider the arithmetic operations in the polar representation of complex numbers. Suppose that z_1 and z_2 are given in the form $z_1 = r_1\big(\cos\varphi_1 + \mathrm{i}\,\sin\varphi_1\big)$ and $z_2 = r_2\big(\cos\varphi_2 + \mathrm{i}\,\sin\varphi_2\big)$. A short computation involving the addition theorems

$$\sin(\alpha + \beta) = \sin\alpha \cdot \cos\beta + \cos\alpha \cdot \sin\beta \tag{2.8}$$

$$\cos(\alpha + \beta) = \cos\alpha \cdot \cos\beta - \sin\alpha \cdot \sin\beta \tag{2.9}$$

yields

$$z_1\,z_2 = r_1\,r_2\,\big(\cos(\varphi_1 + \varphi_2) + \mathrm{i}\,\sin(\varphi_1 + \varphi_2)\big). \tag{2.10}$$

So the *modulus* of a product is the *product* of the moduli of its factors, while the *argument* of a product is the *sum* of the arguments of its factors (as long as one has the correct interpretation of the latter statement).

The absolute value of $z - a$ is the Euclidean distance between the points z and a, which is often used in representations of geometric objects. For example, a *disk* and a *circle* with center a and radius r can be described as

$$D_r(a) := \{z \in \mathbb{C} : \ |z - a| < r\}, \quad T_r(a) := \{z \in \mathbb{C} : \ |z - a| = r\},$$

respectively. The special notation $\mathbb{D}$ and $\mathbb{T}$ is reserved for the *unit disk* and the *unit circle*

$$\mathbb{D} := \{z \in \mathbb{C} : \ |z| < 1\}, \quad \mathbb{T} := \{z \in \mathbb{C} : \ |z| = 1\}.$$

The following arithmetic properties of the modulus can easily be checked,

$$|z_1 z_2| = |z_1| \cdot |z_2|, \quad |z_1/z_2| = |z_1|/|z_2|, \quad |\overline{z}| = |z|, \quad z\,\overline{z} = |z|^2. \tag{2.11}$$

Moreover, we have the estimates

$$\operatorname{Re} z \le |z|, \quad \operatorname{Im} z \le |z|, \quad |z_1 \pm z_2| \le |z_1| + |z_2|.$$

The first two relations follow immediately from the definition of the absolute value. Observing that $|z_1|$, $|z_2|$ and $|z_1 - z_2|$ are the lengths of the sides of the triangle with vertices $0, z_1$ and z_2 in the complex plane, the third inequality (with minus sign) is the *triangle inequality* of plane geometry.

Complex numbers with $|z| = 1$ are termed *unimodular*. In particular the phase $\psi(z) := z/|z|$ of $z \in \mathbb{C} \setminus \{0\}$ is unimodular. Since the sum of the arguments of two factors is an argument of their product, it follows from (2.11) that the phase is multiplicative: for $z_1, z_2 \in \mathbb{C}$ with $z_1, z_2 \ne 0$,

$$\psi(z_1 z_2) = \psi(z_1) \cdot \psi(z_2), \quad \psi(z_1/z_2) = \psi(z_1)/\psi(z_2). \tag{2.12}$$

In particular we have for $z \ne 0$

$$\psi(-z) = -\psi(z), \qquad \psi(1/z) = 1/\psi(z) = \overline{\psi(z)}. \tag{2.13}$$

Powers and Roots. Applying (2.10) repeatedly with one and the same factor z, we get the celebrated De Moivre's formula for the n-th *power* of $z = r\,\big(\cos\varphi + \mathrm{i}\,\sin\varphi\big)$,

$$z^n = r^n\,\big(\cos n\varphi + \mathrm{i}\,\sin n\varphi\big). \tag{2.14}$$

It is easily seen that this formula remains valid if n is a negative integer. In particular, the *inverse* z^{-1} of a complex number $z = r\,(\cos\varphi + \mathrm{i}\,\sin\varphi) \in \mathbb{C} \setminus \{0\}$ can be expressed as

$$z^{-1} = 1/z = r^{-1}\,\Big(\cos(-\varphi) + \mathrm{i}\,\sin(-\varphi)\Big) = r^{-1}\,(\cos\varphi - \mathrm{i}\,\sin\varphi).$$

One main advantage of complex numbers over the reals is that negative numbers have a square root. More generally, a number z is called a *n-th root* of w if it satisfies $z^n = w$. Writing both numbers in polar form,

$$z = r\,\big(\cos\varphi + \mathrm{i}\,\sin\varphi\big), \qquad w = R\,\big(\cos\Phi + \mathrm{i}\,\sin\Phi\big),$$

by (2.7), $z^n = w$ is equivalent to $r^n = R$ and $n\,\varphi = \Phi + 2k\,\pi$ with $k \in \mathbb{Z}$. Resolving this with respect to r and φ yields

$$r = \sqrt[n]{R}, \qquad \varphi = \frac{\Phi}{n} + \frac{2k\,\pi}{n}.$$

Here k is an arbitrary integer, but since adding a multiple of n to k does not change the value of z, it suffices to take the values $k = 0, 1, \ldots, n-1$. So any complex number w different from zero has exactly n roots of order n. In the complex plane, the n-th roots of a non-zero complex number form the vertices of a regular n-gon centered at the origin, as Figure 2.4 shows for the 7-th roots of 1.

If n is a natural number, the *n-th roots of unity* are the solutions of $z^n = 1$. Introducing the *primitive root* of order n,

$$\omega := \cos\frac{2\pi}{n} + \mathrm{i}\,\sin\frac{2\pi}{n},$$

the n-th roots of unity are just the powers $1 = \omega^0, \omega^1, \omega^2, \ldots, \omega^{n-1}$ of ω.

The Point at Infinity. Having solved the problem of extracting roots in the system of complex numbers, we we are still left with another annoying problem with real arithmetic – that of not being able to divide by zero. In order to resolve this, we extend the complex plane by an ideal element, the *point at infinity*, which is denoted by ∞ and is supposed to satisfy

$$\frac{1}{0} := \infty, \qquad \frac{1}{\infty} := 0.$$

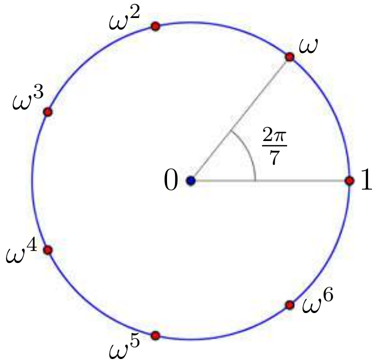


Figure 2.4: The seventh roots of unity

The name is motivated by the observation that the point $1/z$ moves further and further away (to "infinity") as z approaches zero (and vice versa). The union of the complex plane $\mathbb{C}$ with the point at infinity is called the *extended complex plane* and is denoted by $\widehat{\mathbb{C}}$.

The Riemann Sphere. The point at infinity is by no means a mysterious object. It has a simple and beautiful geometric interpretation if complex numbers are considered as points on a *sphere*. The story begins about 2000 years ago, when the Greek mathematician and philosopher Ptolemy invented a method for depicting points from a sphere (in particular from the "celestial sphere") on a flat map. In the 19th century, Bernhard Riemann proposed to utilize *stereographic projection* in the reverse direction for representing complex numbers.

In order to describe this construction, imagine that $\mathbb{C}$ is the XY-plane in three-dimensional XYZ space. To fix the meaning of "up" and "down" we assume that $\mathbb{C}$ is seen in its usual orientation if we look at it "downward from above" (see Figure 2.5).

Let $\mathbb{S}$ be a sphere with radius 1 centered at the common origin of $\mathbb{C}$ and $\mathbb{R}^3$. Taking recourse to concepts of geography, the unit circle $\mathbb{T}$ of the complex plane is the *equator* E of $\mathbb{S}$, and the two points on $\mathbb{S}$ at maximal distance from $\mathbb{C}$ are the *north pole* N (above $\mathbb{C}$) and the *south pole* S (below $\mathbb{C}$).

The *stereographic image* z of a point P on the sphere $\mathbb{S}$ is the intersection of $\mathbb{C}$ with the straight line through P and the north pole N.[2] The point z is well defined for all P on $\mathbb{S}$, with only one exception: the north pole N. If P approaches N, then the distance of the corresponding point z in the plane to the origin gets arbitrarily large. This observation shows that the north pole on $\mathbb{S}$ plays the same role as the point at infinity with respect to the complex plane.

---

[2] Note that there is an alternative definition, where the sphere "lies on top" of the complex plane, touching $\mathbb{C}$ with its south pole.

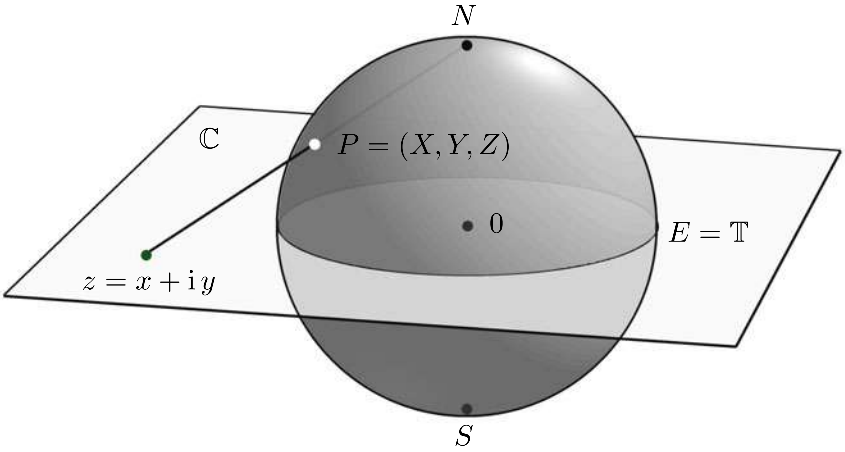


Figure 2.5: Stereographic projection of a sphere onto the complex plane

Extending the stereographic projection to all of $\mathbb{S}$ by assigning the north pole N to the point at infinity results in a bijective correspondence between $\mathbb{S}$ and $\widehat{\mathbb{C}}$. Hence we can label the points on $\mathbb{S}$ with the corresponding complex numbers of $\widehat{\mathbb{C}}$. In what follows we shall therefore identify the sphere $\mathbb{S}$ and the extended complex plane $\widehat{\mathbb{C}}$ and call it the *Riemann sphere*. The *spherical distance* $d(z_1, z_2)$ of two points in $\widehat{\mathbb{C}}$ is the Euclidean length of the straight segment connecting the corresponding points on the sphere $\mathbb{S}$. If $z_1, z_2 \in \mathbb{C}$ then

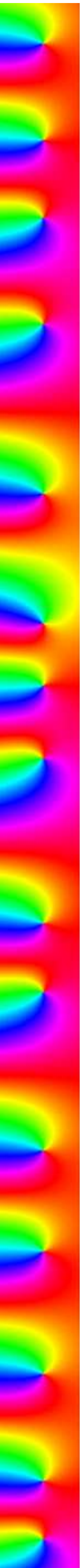
$$d(z_1, z_2) = \frac{2\,|z_1 - z_2|}{\sqrt{1 + |z_1|^2}\sqrt{1 + |z_2|^2}}, \qquad d(z_1, \infty) = \frac{2}{\sqrt{1 + |z_1|^2}}.$$

Arithmetic on the Sphere. Stereographic projection allows us to transplant the arithmetic operations from $\mathbb{C}$ to $\widehat{\mathbb{C}}$. Additionally we postulate that

$$\begin{array}{lll} z/\infty := 0 & \text{for} & z \in \mathbb{C} \\ z/0 := \infty & \text{for} & z \in \mathbb{C} \setminus \{0\} \\ z \pm \infty = \infty \pm z := \infty & \text{for} & z \in \mathbb{C} \\ z \cdot \infty = \infty \cdot z := \infty & \text{for} & z \in \widehat{\mathbb{C}} \setminus \{0\}. \end{array} \tag{2.15}$$

Note that we do not define $\infty \pm \infty$, ∞/∞, $0/0$ and $0 \cdot \infty$. After extending modulus $|z|$ and phase $\psi(z)$ to all points of the Riemann sphere by setting

$$|\infty| := \infty, \quad \psi(0) := 0, \quad \psi(\infty) := \infty,$$



their ranges now become the *extended positive real line* and the *extended unit circle*, respectively, which we define by

$$\widehat{\mathbb{R}}_+ := \mathbb{R}_+ \cup \{0, \infty\}, \qquad \widehat{\mathbb{T}} := \mathbb{T} \cup \{0, \infty\}.$$

Properties of Stereographic Projection. The following properties of stereographic projection are immediate:

(i) The points on the equator remain fixed.

(ii) The lower hemisphere is mapped to the unit disk $\mathbb{D}$.

(iii) The upper hemisphere is mapped to $\mathbb{E}$, the exterior of the unit circle.

Here the exterior of the unit circle is defined such that it includes the point at infinity,

$$\mathbb{E} := \{z \in \mathbb{C} : |z| > 1\} \cup \{\infty\}.$$

Stereographic projection has a remarkable property: it maps circles to circles. More precisely, the image of a circle on $\mathbb{S}$ through N is a line in $\widehat{\mathbb{C}}$, i.e., the union of a line in $\mathbb{C}$ with the point at infinity, while the images of all other circles are proper circles. So the above statement is literally true if we adopt the folkloristic aphorism “lines are circles with infinite radius”.

Computing stereographic projections requires explicit formulas. For this end we represent points z in the complex plane by Cartesian coordinates x, y, and points P on the sphere $\mathbb{S}$ by Cartesian coordinates X, Y, Z in three-dimensional space. The systems are positioned such that the spatial X and Y axes coincide with the x and y axes in the complex plane. Then the coordinates (X, Y, Z) of a point P on $\mathbb{S} \setminus N$ and its stereographic projection $z = x + \mathrm{i}\, y$ are related by the equations

$$x = \frac{X}{1 - Z}, \quad y = \frac{Y}{1 - Z}. \tag{2.16}$$

The north pole $(X, Y, Z) = (0, 0, 1)$ corresponds to the point at infinity, which has no representation in the xy-system. Conversely, if $z = x + \mathrm{i}\, y \in \mathbb{C}$, then

$$X = \frac{2x}{x^2 + y^2 + 1}, \quad Y = \frac{2y}{x^2 + y^2 + 1}, \quad Z = \frac{x^2 + y^2 - 1}{x^2 + y^2 + 1}, \tag{2.17}$$

and $(X, Y, Z) = (0, 0, 1)$ for $z = \infty$.

In summary, we have four interpretations of a complex number $z = x + \mathrm{i}\, y$: as a pair (x, y) of real numbers, as a vector with components x and y, as a point in the Gaussian plane with Cartesian coordinates x and y, and as a point on the Riemann sphere with Cartesian coordinates X, Y, Z given by (2.17).

2.2 Functions and Mappings

A function is a "rule of correspondence" which assigns to each input element a well-defined output element. More precisely, if X and Y are arbitrary sets, a function f from X to Y assigns to any *argument* x in X exactly one *value* $f(x)$ in Y. This is expressed symbolically by writing

$$f : X \to Y,\ x \mapsto f(x).$$

The set X is said to be the *domain set* (or simply the *domain*), Y is called the *target set.*

Mappings or *transformations* are just synonyms for functions. Though there is no basic difference between these concepts, they emphasize different aspects of a function, and the context determines which name is preferred.

It is important to distinguish between a function and its values – while the symbol f stands for the (complete) function, $f(x)$ refers to its value at x. Having mentioned this, we shall nevertheless sometimes say "the function $f(x) = \sin x$", whenever this is convenient and causes no confusion. Occasionally we also write $f(\cdot)$, where the dot is a placeholder for the variable x.

Mapping Properties. The *image* $f(A)$ of a subset A of X is the set of values that f attains on A, and the *pre-image* of $B \subset Y$ is the set of all x in X which are mapped into B,

$$f(A) := \{f(x) \in Y : x \in A\}, \qquad f^{-1}(B) := \{x \in X : f(x) \in B\}.$$

The image $f(X)$ of the domain set is called the *range* of f. A function $f : X \to Y$ is said to be *surjective* if $f(X) = Y$, it is called *injective* (or *one-to-one*) if the equality $f(x_1) = f(x_2)$ only holds for $x_1 = x_2$. A function which is both injective and surjective is termed *bijective.* If f is bijective, there is a unique function $f^{-1} : Y \to X$, the *inverse* of f, which is defined by $f^{-1}(y) = x$ when $f(x) = y$.

A function $f : D \subset \mathbb{C} \to \widehat{\mathbb{C}}$ is said to be *periodic* if there exists a complex number $p \neq 0$ such that for all $z \in D$ also $z + p \in D$ and $f(z + p) = f(z)$. Any such p is called a *period* of f. A period is *primitive* if there is no integer $n > 1$ such that p/n is also a period.

Operations with Functions. Two functions f and g are *equal* if they have the same domain set X and $f(x) = g(x)$ for all $x \in X$. If f and g are defined on X and Y, respectively, with $X \subset Y$ and $f(x) = g(x)$ for all $x \in X$, then f is the *restriction* of g to X and g is an *extension* of f to Y. We shall often use the same symbol to denote a function and its restriction to some subset.

The *composition* of two functions $f : X \to Y$ and $g : Y \to Z$ is the mapping

$$g \circ f : X \to Z,\ z \mapsto g(f(z)).$$

In this book we mainly consider complex functions $f : D \subset \mathbb{C} \to \mathbb{C}$, where the domain set D is a subset of the complex plane $\mathbb{C}$ and the target set is the complex plane, but, more generally, we shall also encounter functions defined on (subsets of) the Riemann sphere $\widehat{\mathbb{C}}$ with values in $\widehat{\mathbb{C}}$. For such functions we denote points $z_0 \in \widehat{\mathbb{C}}$ where $f(z_0) = 0$ or $f(z_0) = \infty$ as *zeros* and *poles* of f, respectively.

If $f : D_f \to \mathbb{C}$ and $g : D_g \to \mathbb{C}$ are complex functions, their sum $f + g$ is defined on $D := D_f \cap D_g$ by $(f+g)(z) := f(z)+g(z)$. A similar definition is made for the difference $f - g$ and the product $f\,g$, while the quotient f/g is naturally defined only on $D \setminus \{z \in D_g : g(z) = 0\}$. If the target set includes the point at infinity, these definitions can be modified accordingly, taking into account the arithmetic rules (2.15).

Functions as Mappings. Figure 2.6 visualizes the action of a complex function as a mapping from a subset of the z-plane to the w-plane. The light yellow regions are the domain set and the range of the function, respectively. Any point z of the domain set is mapped to the corresponding point $f(z)$ in the range. In this manner the function maps ("transplants") the colored objects from the domain to the range.

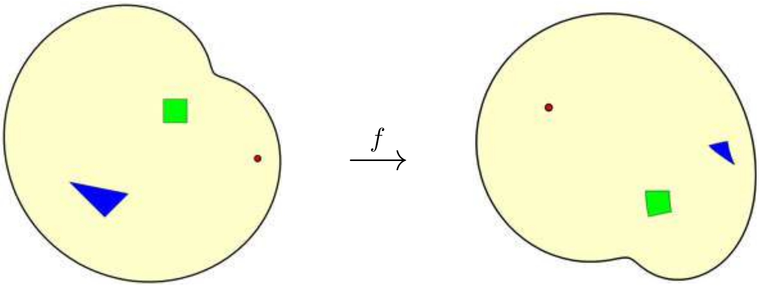


Figure 2.6: A complex function as a mapping which "transplants" sets

In complex analysis the notion of *domain* has two different meanings. The first one alludes to the domain set of a function, while the second pertains to any *open* and *connected* subset of the complex plane or the Riemann sphere (precise definitions of these notions will be given in Section 2.7). Most domain sets of complex functions we shall encounter in this book will indeed be domains in the topological sense.

Decompositions of Complex Functions. A complex function f is composed of its *real part* $u = \operatorname{Re} f$ and its *imaginary part* $v = \operatorname{Im} f$ as $f = u + \mathrm{i}\, v$. Note that $\operatorname{Re} f$ and $\operatorname{Im} f$ are real-valued functions, having the same domain as f. Similarly, f admits a multiplicative decomposition $f = |f| \cdot \psi(f)$ into its *modulus* $|f|$ and its *phase* $\psi(f)$. Unlike the argument $\arg f$, the phase of f is a well-defined function $\psi(f) : D \to \widehat{\mathbb{T}} := \mathbb{T} \cup \{0, \infty\}$.

By a slight abuse of notation, we consider functions f of a complex variable z also as a function of two real variables $x = \operatorname{Re} z$ and $y = \operatorname{Im} z$, thus writing, for example,

$$f(x + \mathrm{i}y) = f(x, y) = u(x, y) + \mathrm{i}\, v(x, y). \tag{2.18}$$

Conversely, if $f : D \subset \mathbb{R}^2 \to \mathbb{R}^2$ is a function of two variables x and y with two real-valued components u and v, then the right-hand side of (2.18) defines a complex-valued function of $z = x + \mathrm{i}\,y$. Using $x = (z + \overline{z})/2$ and $y = (z - \overline{z})/(2\mathrm{i})$, we can rewrite this in the form

$$f(z) = u\left(\frac{z+\overline{z}}{2}, \frac{z-\overline{z}}{2\mathrm{i}}\right) + \mathrm{i}\,v\left(\frac{z+\overline{z}}{2}, \frac{z-\overline{z}}{2\mathrm{i}}\right).$$

Since the right-hand side is an expression in z and $\overline{z}$, some authors prefer to write $f(z, \overline{z})$ instead of $f(z)$. We shall not use this notation since it suggests, rather inappropriately, that the variables $\overline{z}$ and z are independent.

2.3 Arithmetic and Geometry

As we have seen, complex functions can be interpreted as mappings or transformations of (subsets of) the complex plane. In this section we study the arithmetic operations within this framework. In contrast to Section 2.1, here we do not consider operations with two given numbers, instead we fix just one of them and let the second vary through the complex plane.

Addition and Multiplication. Denoting by a and b two given complex numbers with $b \neq 0$, we study the functions

$$f : \mathbb{C} \to \mathbb{C},\ z \mapsto z + a, \qquad g : \mathbb{C} \to \mathbb{C},\ z \mapsto b\,z, \tag{2.19}$$

generated by addition and multiplication. Interpreting a as a vector, it is immediate from the definition of addition that f is a *translation* (or shift) of the complex plane by the vector a. The mapping g can be more conveniently investigated using the polar representation of complex numbers. Here we distinguish several cases.

If b is a *positive number*, then g does not change the argument (or phase) of z, while the modulus of z is multiplied by b. Consequently, g is a *dilation*. If $b > 1$ it stretches the complex plane, and if $b < 1$ it shrinks it, such that each ray emanating from the origin is mapped onto itself and all distances are multiplied by b.

If b is *unimodular*, that is, if $|b| = 1$, then g does not change the modulus of z, while the argument of z is increased by $\beta := \arg b$ (here one can take any fixed argument of b). Thus g is a *rotation* of the plane about the origin by the angle β.

If $b \neq 0$, with $\beta = \arg b$ and $r = |b|$, then g is the composition of the mappings $z \mapsto (\cos\beta + \mathrm{i}\,\sin\beta)\,z$ and $z \mapsto r\,z$, i.e., a rotation by the angle β about the origin followed by a dilation with the same center and stretching factor r. Notice that performing these operations in reverse order yields the same result. We propose to call such transformations *rotostretch* (for the German “Drehstreckung”).

Note that f and g are orientation-preserving similarity transformations in the language of plane Euclidean geometry. In particular, the images of straight lines are straight lines and the images of circles are circles.

Division and Inversion. Finally, we investigate the quotient mapping $z \mapsto b/z$. Since it is the composition of $z \mapsto 1/z$ and $z \mapsto b\,z$, it suffices to study the transformation

$$h : \mathbb{C} \setminus \{0\} \to \mathbb{C} \setminus \{0\}, \ z \mapsto 1/z. \tag{2.20}$$

We first observe that complex conjugation $\mathbb{C} \to \mathbb{C}$, $z \mapsto \overline{z}$ reflects the point z along the real axis, which is an orientation-reversing similarity transformation. We extend this mapping to the Riemann sphere by setting $\overline{\infty} := \infty$.

Next, the modulus and the phase of $1/\overline{z}$ satisfy the relations $|1/\overline{z}| = 1/|z|$ and $\psi(1/\overline{z}) = \psi(z)$, and hence the points z and $1/\overline{z}$ lie on the same ray emanating from the origin and the product of their distances from the origin is equal to one. This characterizes a transformation which is known as *inversion in the unit circle* or simply *inversion*. It sends circles through the origin to straight lines, all other circles are mapped to proper circles. Note that the center of the image circle is not the inverted center of the original circle.

If the inversion is extended to all of the Riemann sphere such that $0 \mapsto \infty$ and $\infty \mapsto 0$, it maps circles on $\widehat{\mathbb{C}}$ to circles on $\widehat{\mathbb{C}}$.
Rewriting the function h in the form $h(z) = 1/z = \overline{1/\overline{z}}$ shows that h is an inversion followed by a reflection across the real axis (or vice versa), which we call an *anti-inversion*.

After extending the functions f, g, and h, defined in (2.19) and (2.20), respectively, to the Riemann sphere by setting

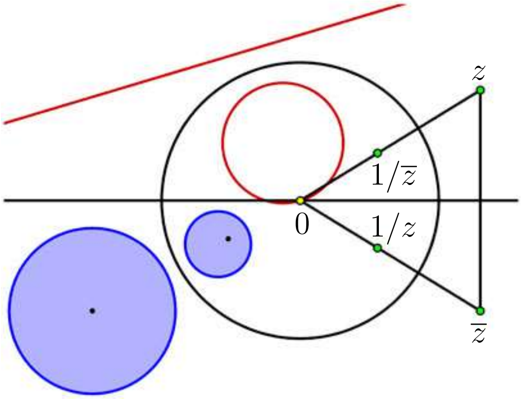


Figure 2.7: Inversion and mapping $z \mapsto 1/z$

$$f(\infty) := \infty, \quad g(\infty) := \infty, \quad h(\infty) := 0, \quad h(0) := \infty,$$

all three functions become bijective mappings of $\widehat{\mathbb{C}}$ onto itself.

Möbius Transformations. It is not difficult to see that all possible compositions of functions f (translation), g (rotostretch) and h (anti-inversion), with arbitrarily chosen values of their parameters a and $b \neq 0$, have a specific form, namely

$$F(z) = \frac{Az + B}{Cz + D}, \quad \text{with} \quad AC - BD \neq 0.$$

Functions of this type play a prominent role in complex analysis. In honor of August Ferdinand Möbius they are called *Möbius transformations*. All Möbius transformations are bijections of the Riemann sphere onto itself and form a *group* with respect to composition. In particular the inverse of F is

$$F^{-1}(z) = \frac{Dz - B}{-Cz + A}.$$

Conversely, any Möbius transformation can be composed of at most two translations, one rotostretch, and one anti-inversion. For $C = 0$ this is trivial, if $C \neq 0$, then $w = F(z)$ is a composition of

$$z \mapsto w_1 := \frac{C^2}{BC - AD}\, z \mapsto w_2 := w_1 + \frac{CD}{BC - AD} \mapsto w_3 := \frac{1}{w_2} \mapsto w := w_3 + \frac{A}{C}.$$

The investigation of Möbius transformations will be continued in Section 6.3 which is exclusively devoted to this class of functions.

2.4 The Analytic Landscape

By now we already know some special complex functions, and it might be interesting to *see* them. Real functions can be conveniently depicted by their *graph.* But when we try to do the same for a complex function $f : D \subset \mathbb{C} \to \mathbb{C}$ we are quickly stumped because the graph

$$G_f := \big\{ \big(z, f(z)\big) \in \mathbb{C} \times \mathbb{C} : z \in D \big\}$$

of f lives in $\mathbb{C} \times \mathbb{C}$, which has four real dimensions. In order to stay in three spatial dimensions, we need a substitute for the missing fourth dimension.

As has already been discussed in Chapter 1, one option is to start with the traditional analytic landscapes and to incorporate the missing information by color.

The *analytic landscape* A_f (also known as the *relief* or *module surface*) of a function $f : D \subset \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ is the graph of its absolute value,

$$A_f := \big\{ \big(z, |f(z)|\big) \in \widehat{\mathbb{C}} \times \widehat{\mathbb{R}}_+ : z \in D \big\}.$$

Analytic landscapes involve only the *modulus* of a function and leave out its *argument.* In order to avoid the ambiguity of the latter, we shall work with the *phase* instead. Since the phase of non-zero complex numbers lives on the unit circle $\mathbb{T}$, and points on a circle can naturally be encoded by colors, color is an ideal candidate for visualizing phase.

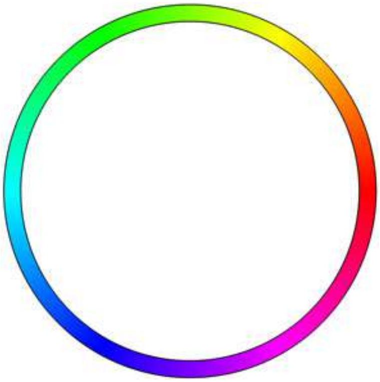


Figure 2.8: The color circle

Colored Analytic Landscapes. The *colored analytic landscape* of a complex function is the graph of its modulus colored according to its phase. Formally, the colored analytic landscape C_f of a function $f : D \to \widehat{\mathbb{C}}$ can be defined as the set

$$C_f := \big\{ \big(z, |f(z)|, \psi(f(z))\big) \in \widehat{\mathbb{C}} \times \widehat{\mathbb{R}}_+ \times \widehat{\mathbb{T}} : z \in D \big\},$$

where z is the position of the base point, $|f(z)|$ is interpreted as the height of the point, and the phase $\psi(f(z))$ determines its color.

The picture on the left in Figure 2.9 shows the colored analytic landscape of the function

$$f(z) = (z-1)/(z^2+z+1) \tag{2.21}$$

in the square $|\operatorname{Re} z| \leq 2$, $|\operatorname{Im} z| \leq 2$, which will henceforth serve as a *standard example*. The function has a *zero* at $z_0 = 1$, i.e., $f(z_0) = 0$. At $z_1 := (-1+\sqrt{3}\,\mathrm{i})/2$ and $z_2 := (-1-\sqrt{3}\,\mathrm{i})/2$ the denominator of $f(z)$ vanishes. We set $f\big(z_{1/2}\big) := \infty$ and refer to z_1 and z_2 as the *poles* of f.

If the modulus of a function varies over a wide range, it is better to use a *logarithmic scaling* of the vertical axis. This representation is also more natural since $\log|f|$ and $\arg f$ are conjugate harmonic functions (see Section 4.6). The corresponding *colored logarithmic analytic landscape* of the standard example is depicted in Figure 2.9 (right).

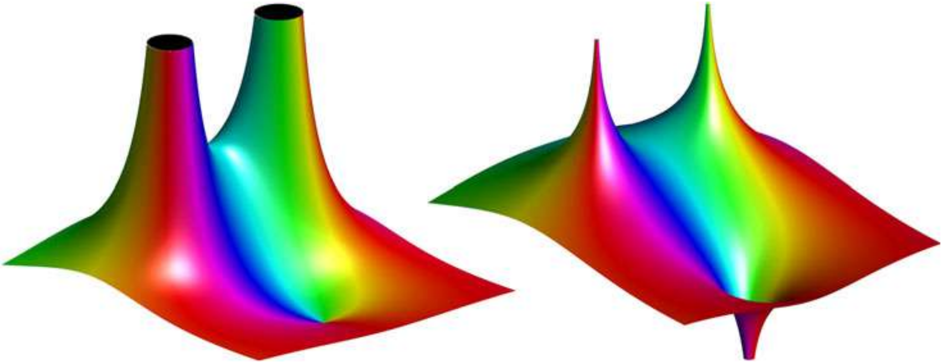


Figure 2.9: Colored analytic landscapes with conventional and logarithmic scaling

The Color Scheme for Phase. Of course the color coding of the phase is by no means unique. It is clear that saturated colors are preferable because they can be identified and distinguished better. This, in some sense, picks out the 'hue' component of the `HSV` (hue, saturation, value) color scheme. The corresponding color wheel is shown in Figure 2.8. For psychological reasons, the standardized color circle has been rotated such that positive values are encoded red. Finally, the exceptional values zero and infinity are associated with black and white, respectively. This color scheme will be used throughout the text. We would like to encourage others to adopt the same color scheme to make phase portraits comparable.

2.5 Color Representations

With colored analytic landscapes the problem of visualizing complex functions could be considered solved. However, the module surface is a three-dimensional object which usually must be projected onto a two-dimensional sheet of paper or a screen for visualization. Often this projection causes problems since interesting elements (like zeros) become invisible or are hard to detect.

Domain Coloring. But there is yet an alternative approach which is not only simpler but also more general. Having introduced colors, it is kind of natural to use them not only for representing the phase of a function, but also to *completely encode its values, using a two-dimensional color scheme.* Then, instead of drawing a graph, one can depict a function directly on its domain by color-coding its values, thus converting it to an *image.*

Coloring techniques have been customary for many decades, for example in depicting altitudes or temperatures on maps, but in most cases they represent *real-valued* functions using a one-dimensional color scheme. Two-dimensional color schemes for visualizing complex valued functions have been in use at least since the late 1980s (Larry Crone [7], see Hans Lundmark [38]), but they became popular only with Frank Farris' review [15] of Tristan Needham's book [44]. Farris also coined the suggestive name *domain coloring.*

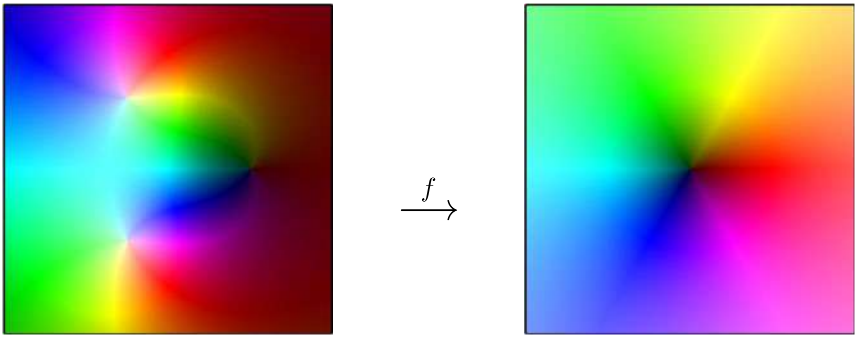


Figure 2.10: Domain coloring of the function $f(z) = (z-1)/(z^2+z+1)$

Figure 2.10 (left) shows a domain coloring for our standard example. The related color scheme for the values in the complex w-plane is shown in the window on the right. As usual, phase is encoded as "color" (in the more precise meaning of *hue*), while "lightness" corresponds to the modulus. Any point z in the domain of f carries the same color like its image $f(z)$ in the w-plane.

Compared to the colored analytic landscape, domain coloring has the advantage that functions are represented in two dimensions, which makes it easier to visualize complicated functions.

Domain coloring pictures are often very beautiful and some are even quite artistic (see Crone [7], or Glaeser and Polthier [18]). On the other hand they may be somewhat fuzzy, which makes it difficult, for example, to locate zeros precisely. Moreover, typically the modulus of a function varies in a wide range, while the human eye normally is not very sensitive to different shades of gray. Thus the color scheme must be appropriately tuned to the function – using one and the same standardized scheme for all functions does not always give satisfactory results.

Phase Portraits. What happens, if we forget about the modulus completely and just depict the color-coded phase? This *phase portrait* is exactly what we see looking at the colored analytic landscape straight from the top in the direction perpendicular to the xy-plane. Figure 2.11 shows the result for the example function defined in (2.21).

To give a formal definition, if f is a complex function on D, then the mapping

$$\Psi_f : D \to \widehat{\mathbb{T}},\ z \mapsto \psi\big(f(z)\big)$$

will be designated as the *phase* of f, and its graph

$$P_f := \{(z, \Psi_f(z)) : z \in D\}$$

is referred to as the *phase portrait* or *phase plot* of f. If $\widehat{\mathbb{T}} := \mathbb{T} \cup \{0, \infty\}$ is identified with the color circle extended by black (corresponding to 0) and white (corresponding to ∞), then the phase portrait of a function can be interpreted as an image.

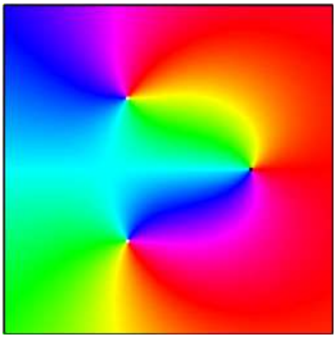


Figure 2.11: A phase portrait

Phase Portraits Versus Analytic Landscapes. In a sense, phase portraits are complementary to the (uncolored) analytic landscape: one neglects modulus, the other one omits phase. So it seems that not much is gained, but indeed there are some essential advantages of phase portraits over analytic landscapes.

First of all, the phase portrait is a two-dimensional image, which does not need to be projected, and our brain is well trained in interpreting such images. Secondly, compared with the range of the modulus of a typical function, the range of the phase is quite small since it is a subset of the extended unit circle. Consequently the visual resolution is much higher for the phase than for the absolute value, which allows us to represent all functions with one and the same color scheme. Thirdly, the reconstruction of missing information is simpler and more accurate for phase plots. We cannot discuss this in detail right now, but, hopefully, it will become clear in due course.

On the other hand, one must also mention that phase portraits are not appropriate to visualize all complex functions because two different functions may have the same phase portrait when they differ only in their modulus.

The situation changes when we restrict ourselves to the important class of *analytic functions* which are of prime importance in this book. As we shall see in Section 3.4 (Corollary 3.4.9), such functions are *completely characterized* (up to a positive scaling factor) by their phase portraits.

Enhanced Phase Portraits. Since phase occupies only one dimension, there is plenty of room in the color space to incorporate additional information. For example, encoding the modulus in a scale of gray would lead us back to conventional domain coloring. But we can make other more interesting modifications as well. Which one we shall choose depends on the properties which we would like to emphasize and perhaps also on the function which we are investigating. In order to understand this better, let us reconsider the phase portrait from a slightly different point of view.

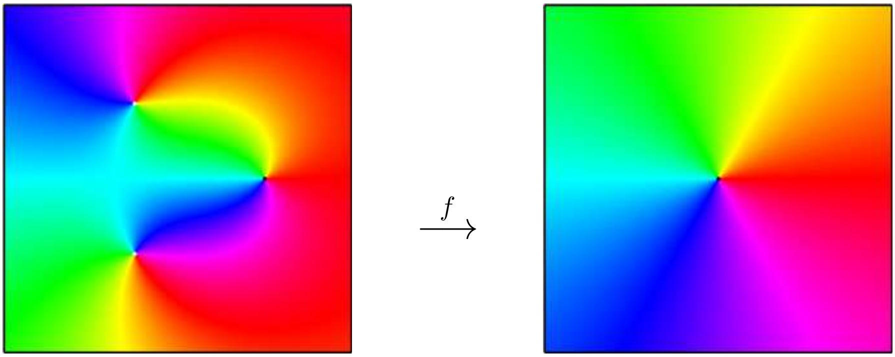


Figure 2.12: The phase portrait as pull back of the colored w-plane

Figure 2.12 shows a function f mapping the complex z-plane to the complex w-plane. In order to generate the phase portrait of f, first the w-plane is colored according to the phase of its points. In the second step, every point z in the domain of definition of f gets the same color as the value $f(z)$ has in the w-plane. In short, the phase portrait on the left is the *pull back* of the picture on the right by the function f.

Let us pause here for a moment and think about what could happen if we would try to transplant an image in the other direction. *Pushing forward* an image from the z-plane to the w-plane via f means that the color of every point z in the domain D is transplanted to the corresponding point $f(z)$ in the range of f. But this causes a conflict whenever two points z_1 and z_2 are colored differently and f attains the same value at z_1 and z_2.

While pushing forward may be problematic, it is clear that *any* picture in the w-plane can be pulled back to the z-plane by the function f. This provides many options to modify and enhance the color scheme of phase portraits.

In this book we shall mainly use three variations of phase portraits which are depicted on page 32 for the standard example $f(z) := (z-1)/(z^2+z+1)$.

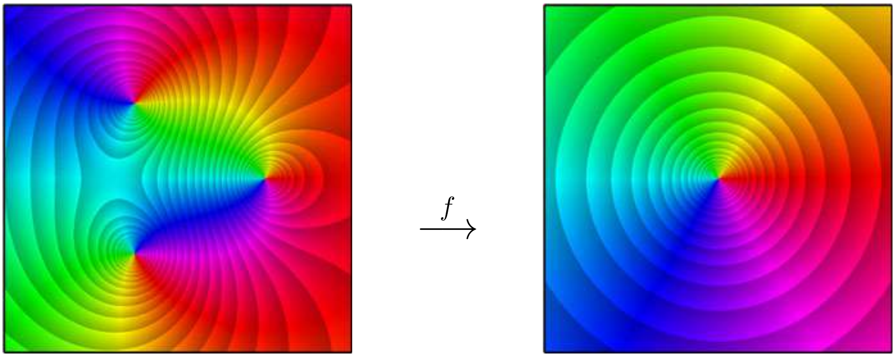


Figure 2.13: Generation of a phase portrait with modulus contour lines

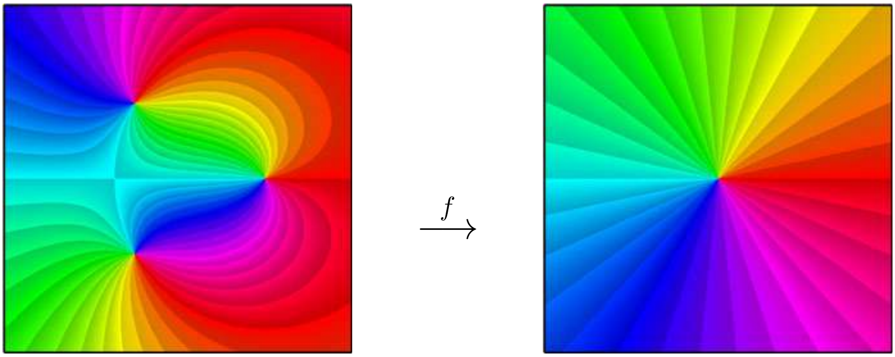


Figure 2.14: Generation of a phase portrait with phase contour lines

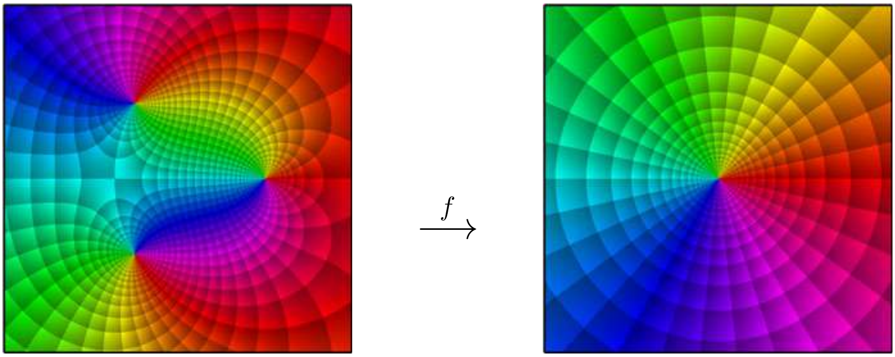


Figure 2.15: A phase portrait with contour lines of modulus and phase

The visible domain is the square given by $|\mathrm{Re}\, z| \le 2$, $|\mathrm{Im}\, z| \le 2$. Recall that the function has a zero at $z_0 = 1$ and two poles at $z_{1/2} = (-1 \pm \sqrt{3}\,\mathrm{i})/2$. The right-hand windows of all three figures show the color scheme of the w-plane, on the left we see the corresponding enhanced phase plots generated by pulling back the right image by the function f.

The pictures in Figure 2.13 involve a gray component g which is a sawtooth function of $\log|f|$, like

$$g = \lceil \log|f| \rceil - \log|f|.$$

Here $x \mapsto \lceil x \rceil$ is the *ceiling function*, which determines the smallest integer not less than x. Figure 2.16 shows the function $x \mapsto \lceil x \rceil - x$ (left) and a typical gray intensity as function of $|f|$ (right).

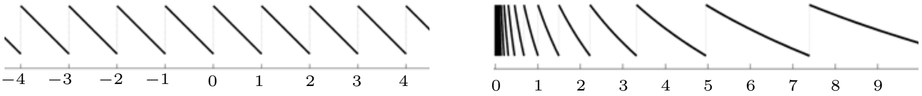


Figure 2.16: The function $x \mapsto \lceil x \rceil - x$ and the gray value as function of $|f|$

The jumps in the gray component generate *contour lines* of $|f|$, i.e., lines of constant modulus. In between two such lines darker colors correspond to smaller values of $|f|$. Though this coloring is relatively insensitive to the range of the function depicted, the distance between adjacent contour lines can be adjusted by modifying the frequency of the sawtooth function g.

It is worth noticing that the contour lines arise from a visual effect and need not really be computed. The proposed shading methods are stable, they do not require sophisticated numerical algorithms, and work with almost no additional computational effort.

Figure 2.14 demonstrates an analogous effect for the phase. In the plain phase portrait the sets of constant phase are *isochromatic*. Here some of these lines are enhanced by the discontinuities of the shading.

In Figure 2.15, the gray value is the product of two sawtooth functions depending on the logarithm of the modulus and the phase of w, respectively. In the w-plane the discontinuities of this shading generate a (logarithmically scaled) polar *tiling*.

Notice that the frequencies of the sawtooth functions encoding modulus and phase are not independent of each other, but chosen such that the tiles are “almost squares”. This vague statement will be made precise in Section 6.1, it suffices here to say that all tiles have the property that they have four right-angled corners and four sides of approximately the same lengths.

The corresponding tiles in the z-plane, generated by the pull-back via the function f, are shown in Figure 2.15 (left). Somewhat surprisingly, these tiles also seem to possess the same property, as far as visual inspection allows us to decide and provided that we ignore some exceptions. The reason behind this observation is a specific property of the mapping $z \mapsto f(z)$: the transplantation via f preserves the angle of intersection between curves. Mappings with this property are called *angle-preserving* or *conformal* and will be studied in Section 6.1.

Other Color Schemes. While investigating conformal maps, functions will be visualized using black-and-white coloring, because it depicts conformality more clearly and suppresses irrelevant information. Two such schemes can be seen at the top and in the middle of page 35.

The starting point of Figure 2.17 is a polar chessboard-like tiling of the w-plane. The corresponding coloring of the domain reflects the phase and the (logarithm of the) modulus of f. This figure is a black-and-white version of Figure 2.15. Observe the high resolution at the three points where f is zero or infinity.

In Figure 2.18, the w-plane is colored in the conventional Cartesian chessboard style. In its pull-back to the z-plane we see an extremely fine structure near the poles of f, where a large number of squares from the w-plane is compressed into a small region. On the other hand, the zero of f cannot be seen at all, as its neighborhood does not have any special structure.

The last figure neglects phase information completely. The ring-shaped alternating black and white stripes in Figure 2.19 are the sets

$$\{z \in D: \ k\,d < \log|f(z)| \le (k+1)\,d\}, \quad \{w \in \mathbb{C}: \ k\,d < \log|w| \le (k+1)\,d\}$$

with $k \in \mathbb{Z}$ and some positive d. In the left picture, their boundaries are contour lines of f. This type of coloring is particularly useful in producing equipotential lines in plane electrostatics, which will be demonstrated in Section 4.6.

The color schemes which are based on a polar grid in the w-plane yield a high resolution near zeros (and poles) of f in the z-plane. This ‘microscope effect’ can be utilized to explore the structure of a function f in a neighborhood of any other point of its domain as well. To enhance the resolution at a point a in the z-plane, just shift the origin of the polar grid in the w-plane to the point $b = f(a)$. The same effect is achieved by considering the function $f(z) - f(a)$ in the usual scheme. One should be aware that this transformation changes phase and modulus of the function, and that there is no simple relation between the phases of the functions f and $f - a$.

Finally, we mention that the images on the left side of all figures are constructed as the pull-back of the *entire* colored complex w-plane, while the windows on the right-hand side show only a section illustrating the color scheme.

Since it is the declared goal of this book to promote phase portraits as a tool for visualizing and exploring complex functions, we refrain from using conventional domain coloring. Though, in principle, we will prefer to explain ideas and concepts using plain phase portraits as much as possible, we shall often resort to modifications if they demonstrate relevant facts more clearly. Another motive for making modifications is aesthetics; for example, when the plain phase portraits lack an interesting structure.

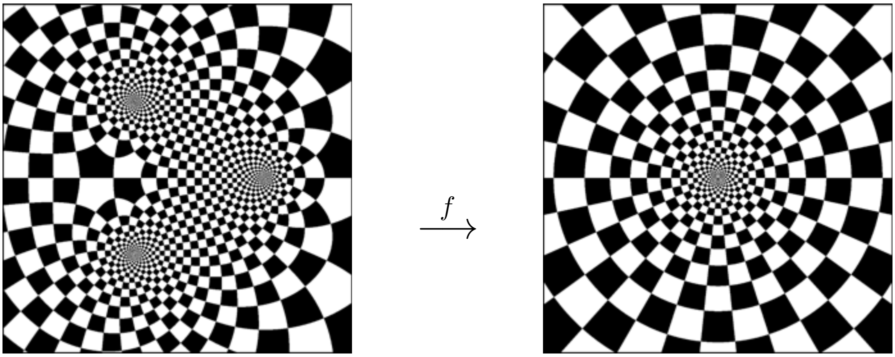


Figure 2.17: The pull back of a polar chessboard visualizes conformality

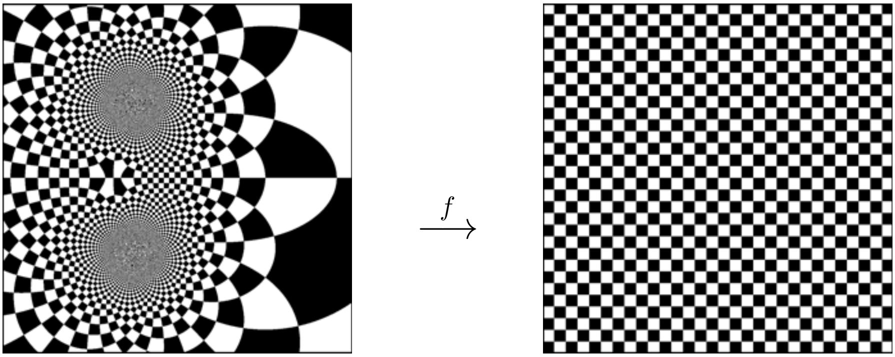


Figure 2.18: The pull back of a Cartesian chessboard hides zeros

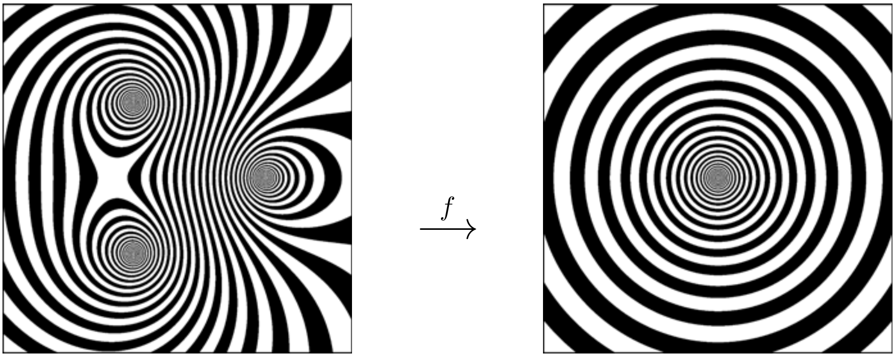


Figure 2.19: The pull back of circular rings generates modulus contour lines

Practical Excursion. With this much preparation, we can now try to get some practice in working with phase portraits. We shall introduce some more basic notions at a somewhat informal level as we go along.

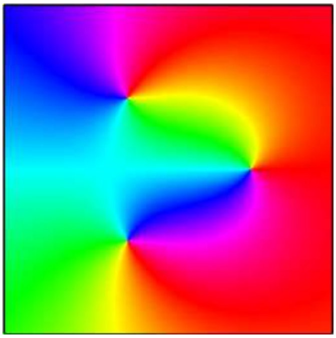


Figure 2.20: A phase portrait

Figure 2.20 depicts again the familiar example $f(z) = (z-1)/(z^2+z+1)$. We have already mentioned that the three exceptional points where all colors come together correspond to the zero and the two poles of the function. Notice that zeros and poles can be distinguished by the ordering of colors in their neighborhood.

A more careful inspection reveals that there might be yet another special point in the light blue region on the negative real axis. Since it is located in a diffused spot, we modify the phase coloring (by overlaying a discontinuous gray component) to get a sharp contrast in the color of interest. Adjusting the jump in the color scheme correctly, we get the result shown on the left in Figure 2.21. Now the exceptional point is clearly visible, it is the crossing point of (two) smooth curves having the same color. We shall call such points *saddles*.

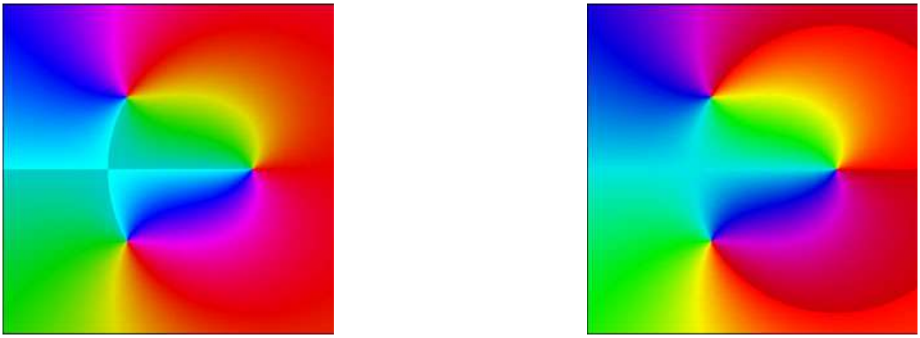


Figure 2.21: Two phase portraits with enhanced isochromatic lines

Isochromatic Sets. In order to describe this more precisely and to explore the structure of phase portraits in some detail, we introduce *isochromatic sets*. If f is a function on a domain D these are defined by

$$S(c) := \{z \in D : \psi(f(z)) = c\}, \quad c \in \widehat{\mathbb{T}}.$$

If the point c on the unit circle is identified with the corresponding color on the color wheel, then $S(c)$ is the set of all points in D carrying the color c.

In Figure 2.21 (left) the blue line and the crossing arc constitute the isochromatic set $S(-1)$. Similarly, in the window on the right the color changes abruptly across three red arcs belonging to the isochromatic set $S(1)$.

Complex and Analytic Functions. If we do not impose additional restrictions, like continuity or differentiability, the isochromatic sets of complex functions can be arbitrary – but this is not so for "analytic" functions, which are the objects of prime interest in this text. Further investigation of this question is deferred to the next chapter and will be continued in more detail in Volume 2 where we shall see that any isochromatic set of an analytic function is the union of smooth arcs which can be linked only at saddle points in a very specific way.[3]

Warning. *Most statements and results in the following chapters are about analytic functions and do not hold for complex functions in general.*

In Section 3.4, we shall see that *analytic functions* are (almost) uniquely determined by their (pure) phase portraits (compare Theorem 3.4.10), but this is not so for general functions. For example, the functions f (analytic) and g (not analytic) defined by

$$f(z) = (z-1)/(z^2+z-1), \qquad g(z) = (z-1)\cdot(\overline{z}^2+\overline{z}-1), \tag{2.22}$$

have the same phase (except at their zeros and poles) though they are completely different.

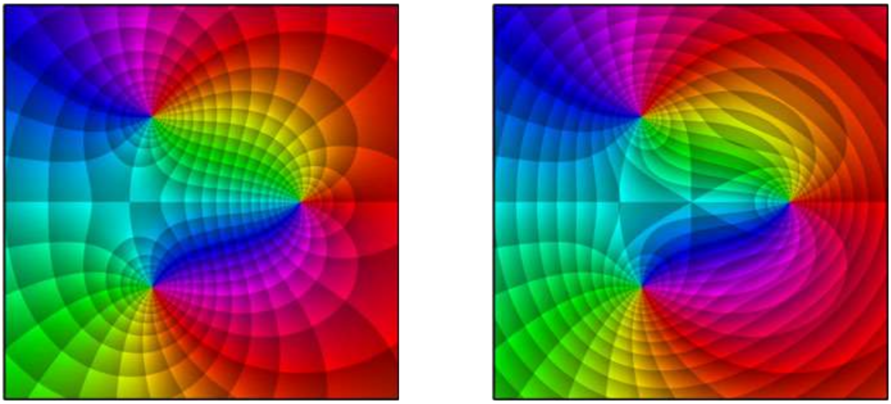


Figure 2.22: Enhanced phase portraits of f (left) and g (right)

Since pure phase portraits do not always display enough information for exploring general complex functions, we recommend use of their enhanced versions with contour lines of modulus and phase in such cases. Figure 2.22 shows two such portraits of the functions f (left) and g (right) defined in (2.22).

---

[3] See also Wegert [69].

A notable distinction between the two portraits is the shape of the tiles. In the left picture most of them are almost squares and have right-angled corners. In contrast, many tiles in the portrait of g are prolate and their angles differ significantly from $\pi/2$ – at some points the contour lines of modulus and phase are even mutually tangent. We shall explore these observations in Section 6.1.

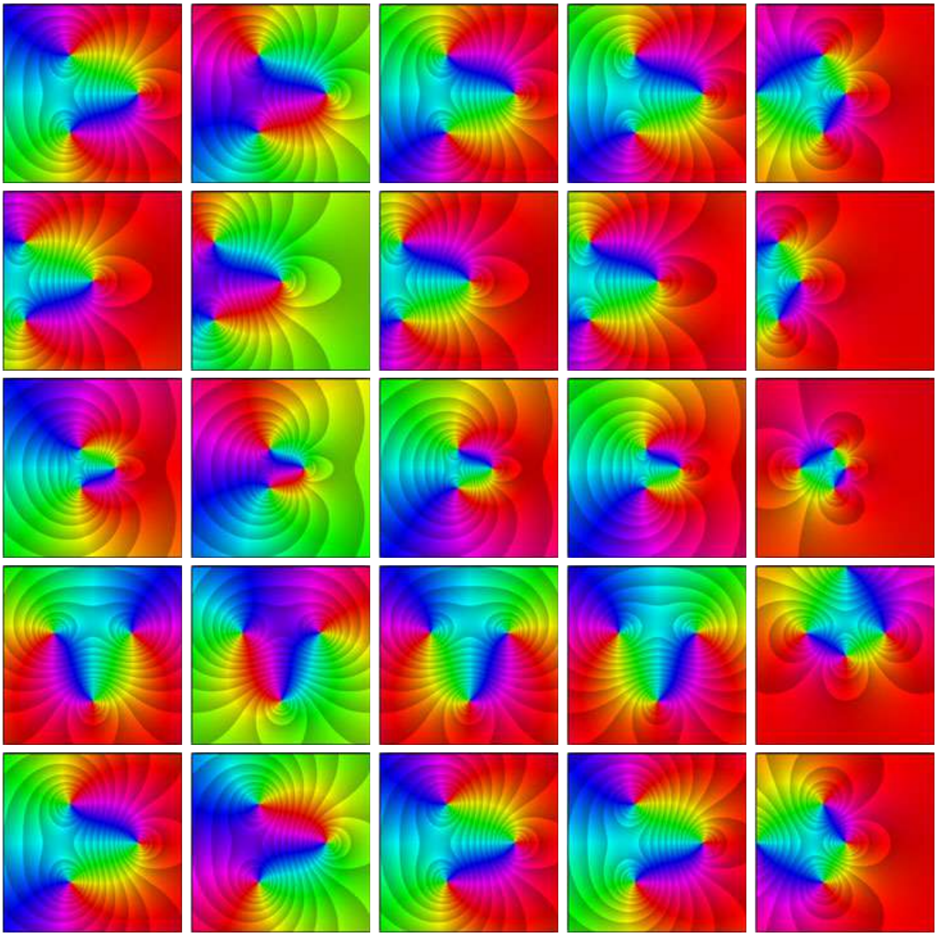


Figure 2.23: Phase portraits with contour lines of compositions $h \circ f \circ g$

The Effect of Transformations. Figure 2.23 demonstrates what happens with the phase portrait of a function $w = f(z)$ when the variables z and w are transformed. It displays the phase portraits of the compositions $h \circ f \circ g$, where g corresponds to the rows and h is associated with the columns, and, in this order,

$$g : z \mapsto z, \quad z+1, \quad 2z, \quad \mathrm{i}\,z, \quad \overline{z}, \qquad h : w \mapsto w, \quad \mathrm{i}\,w, \quad 1/w, \quad \overline{w}, \quad w+1.$$

The domain depicted is the square $-2 \leq \operatorname{Re} z, \operatorname{Im} z \leq 2$. Notice that, compared to the first row, the plots are shifted to the left in the second row, shrunk by a factor of two in the third row, rotated clockwise by $\pi/2$ in the fourth row, and reflected at the real axis in the fifth row.

Though the pure phase portraits of the third and the fourth columns would be identical, there is a difference in the gray shading which represents the modulus. The fifth column seems to have no relation to the others.

Phase Portraits on the Sphere. Phase portraits of functions defined on the extended complex plane can be drawn directly on the Riemann sphere. Though this idea is natural, and results in quite nice pictures (and I cannot resist the temptation of showing you the transplantations of the images on page 35 to the sphere), it has several disadvantages.

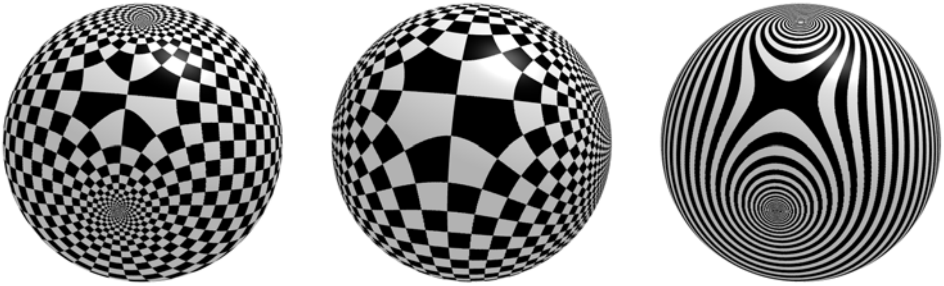


Figure 2.24: A function on the sphere represented by black-and-white schemes

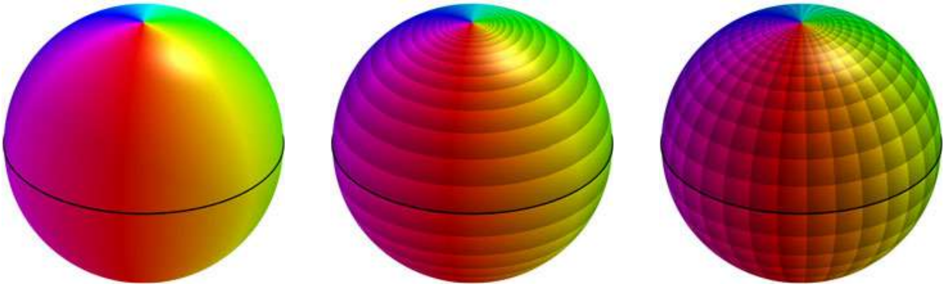


Figure 2.25: Three color schemes on the Riemann sphere

Figure 2.25 shows the coloring of $\mathbb{S}$ according to the three color schemes on page 32, which can be interpreted as the (modified) phase portraits of the function $f(z) = z$.

Since only part of the sphere can be seen from a fixed position, at least two pictures are needed to represent the complete sphere, and due to perspective projection, precise localization of points is difficult. Also, the lighting and shading needed to make the sphere look round distorts the colors. But worst of all, when we inspect a neighborhood of the point at infinity, where the value of the function $f(z) = z$ is infinite, we see a coloring which resembles a zero. Conversely, the neighborhood of a zero is colored in a manner similar to that for poles. Why does this happen?

The reason is that stereographic projection acts between the *upper side* of the plane and the *interior side* of the sphere. This becomes obvious when we consider the projection of the lower hemisphere onto the unit disk, but it remains true for the upper hemisphere as well. Looking at a sphere from the outside, as we usually do, *reverses the orientation*, which makes zeros look like poles and vice versa.

Of course, once we are aware of this phenomenon, we could live with it and need not change anything, except mentally altering orientation whenever we switch between the plane and the sphere. However, since orientation is crucial in the interpretation of phase portraits, this may be rather confusing.

A simple alternative is to represent the sphere $\mathbb{S}$ by *two charts* which depict what we see looking at $\mathbb{S}$ from the inside. The first one is the stereographic projection P_N of the sphere $\mathbb{S}$ from the north pole N (identified with the point at infinity) to a complex plane $\mathbb{C}$ attached to the *upper side* of the horizontal plane P. The projection P_N maps $\mathbb{S} \setminus \{N\}$ onto $\mathbb{C}$ and will be used to represent the lower hemisphere.

The second chart P_S is the stereographic projection of $\mathbb{S}$ from the south pole S to a complex plane $\mathbb{C}$ attached to the *lower side* of P. It maps $\mathbb{S} \setminus \{S\}$ onto $\mathbb{C}$ and will be used to represent the upper hemisphere.

The result of transplanting the phase portrait from $\mathbb{S}$ to the two copies of $\mathbb{C}$ via the charts P_N and P_S is shown in Figure 2.27. In the left window the image of the lower hemisphere is highlighted, while the saturated colors in the right window emphasize the image of the upper hemisphere. In both pictures the images of the points $1, \mathrm{i}, -1, -\mathrm{i}$ are marked.

All points on $\mathbb{S}$, except the two poles N and S, are represented in both charts. This defines a *transition map* between the images of the intersection of the domains of both charts,

$$P_S \circ P_N^{-1} : \ \mathbb{C} \setminus \{0\} \to \mathbb{C} \setminus \{0\}, \ z \mapsto 1/z.$$

Intuitively, the colored sphere can be modelled from the two pictures in Figure 2.27 by the following construction: cut off the two highlighted disks, put the pieces together face-to-face (colors inside) so that the marked points (and then all points on the unit circle) fit and glue them along the unit circle. Now blow it up to a spherical balloon!

After rotating the right window by an angle of π about the origin, as shown in Figure 2.28, this procedure can be simplified: fold at the dashed line, glue along the black circle, cut off the outer part, and inflate it.

Since this interpretation of Figure 2.28 is intuitive, we shall mainly use these pictures for representing colorings of the Riemann sphere.

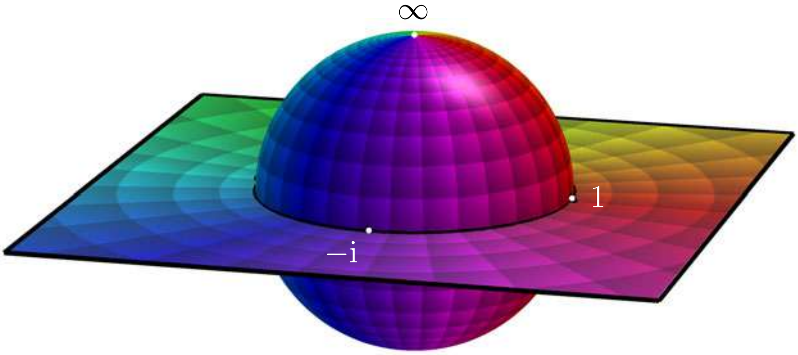


Figure 2.26: Stereographic projection of phase portraits

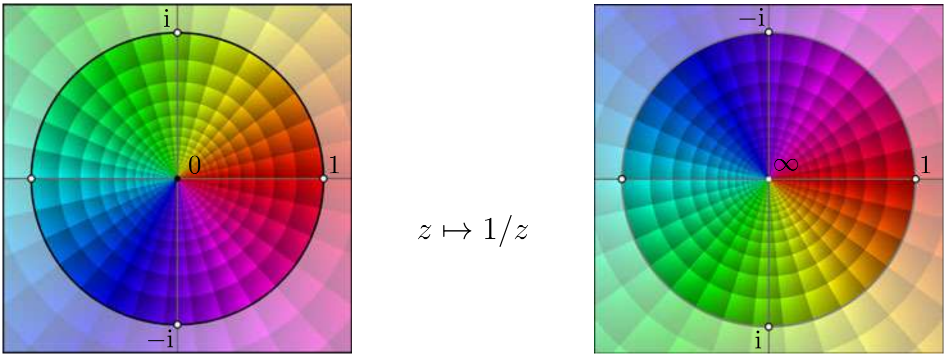


Figure 2.27: Two charts of the sphere corresponding to the hemispheres

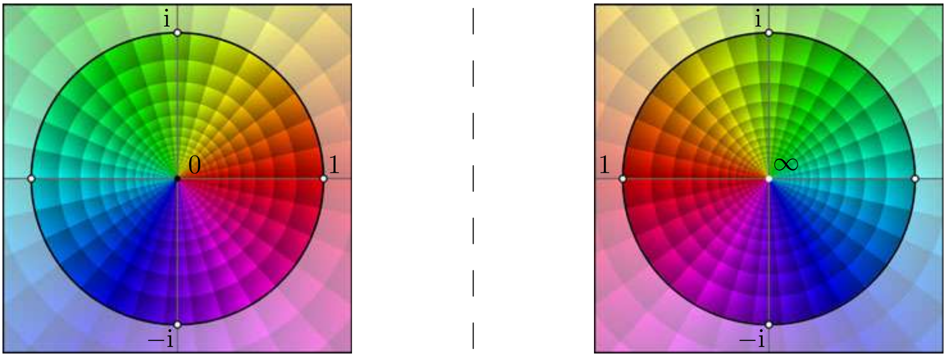


Figure 2.28: The standard representation of the sphere by two charts

An Example. Figure 2.29 depicts an enhanced phase portrait with modulus and phase contour lines of the standard example $f(z) = (z-1)/(z^2+z+1)$ on the Riemann sphere. The picture on the left shows a region which we have already seen before – the domain in the right is the exterior of the unit disk with the point at infinity in the center.

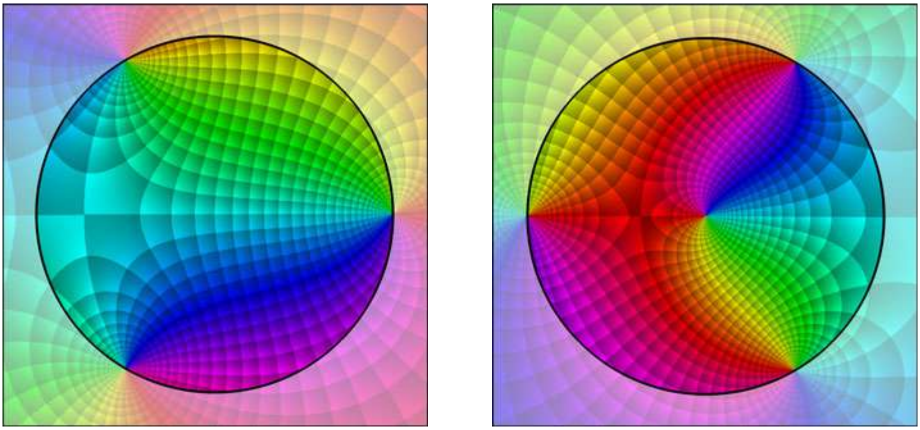


Figure 2.29: Enhanced phase portrait of the standard example on the sphere

It seems that something special happens at this point: it is one of the four distinctive points on the sphere where all colors of the chromatic circle meet. The other three points of this type are the zero and the two poles on the unit circle. Though the function is not yet defined at $z=\infty$, we observe that its phase portrait *in a neighborhood of this point* resembles the typical behavior at a zero. The function somehow seems to *request* a zero at infinity! So we satisfy its desire by setting $f(\infty) := 0$. Why this is indeed the right choice will be investigated in the next section.

2.6 Convergence and Continuity

Though we assume that the reader has some familiarity with these notions, we review some basic facts without proofs.

Sequences. A *sequence* $(z_n)_{n=1}^{\infty}$ in a set A is a function on the set $\mathbb{Z}_+$ of positive integers with values in A. For brevity we shall often use the simplified notation (z_n), but sometimes we also use the longish but more intuitive form $z_1, z_2, z_3, \ldots$. In order to indicate that (z_n) is a sequence in A we write $(z_n) \subset A$. Most sequences we shall meet in this book are sequences in $\mathbb{C}$, or more generally, in $\widehat{\mathbb{C}}$.

Note that the indexing of a sequence need not necessarily start with $n=1$, the point is that the elements of a sequence are strictly ordered, which *in principle* allows indexing by the natural numbers.

The concept of convergent sequences in $\mathbb{C}$ is analogous to that for sequences of real numbers: a sequence $(z_n) \subset \mathbb{C}$ *converges* to $z_0 \in \mathbb{C}$ if [4] for every $\varepsilon > 0$, there is an integer N such that $n \geq N$ implies $|z_n - z_0| < \varepsilon$. If such a number z_0 exists it is unique and is said to be the *limit* of the sequence (z_n). Sequences which do not converge are termed *divergent*. Convergence of (z_n) to z_0 is written as

$$\lim_{n\to\infty} z_n = z_0, \quad \text{or} \quad z_n \to z_0.$$

Less formally, a sequence (z_n) converges to z_0 if the points z_n get arbitrarily close to z_0 as n gets sufficiently large.

Limits of complex sequences have the same arithmetic properties as limits of real sequences, for example $z_n \to z_0$ and $w_n \to w_0$ imply that $z_n + w_n \to z_0 + w_0$.

Convergence of sequences in $\widehat{\mathbb{C}}$ is defined in a similar manner, with only a single modification: the Euclidean distance $|z - w|$ between two points z and w in the plane is replaced by the spherical distance $d(z, w)$. Convergence in $\mathbb{C}$ is compatible with this more general concept; this implies that if $z_0 \in \mathbb{C}$, then the sequence $(z_n) \subset \mathbb{C}$ converges to z_0 in $\mathbb{C}$, if and only if it converges to z_0 in $\widehat{\mathbb{C}}$. Convergence of z_n to ∞ can be rephrased as $1/z_n \to 0$, that is, for every positive ε there exists an N such that $|z_n| > 1/\varepsilon$ for all $n \geq N$.

Continuity. Another fundamental concept of analysis is continuity of functions. Roughly speaking, a function $f : X \to Y$ is continuous at a point $x_0 \in X$ if the distance of $f(x)$ and $f(x_0)$ becomes arbitrarily small whenever x in X is sufficiently close to x_0. Expressed in ε-δ-language, and restricted to complex functions, this reads as follows:

A function $f : D \subset \mathbb{C} \to \mathbb{C}$ is *continuous* at $z_0 \in D$ if for every positive ε there exists a positive δ such that $z \in D$ and $|z - z_0| < \delta$ imply $|f(z) - f(z_0)| < \varepsilon$.

We say that f is continuous on D, if it is continuous at each point of D. Then, in general, the value of δ depends on ε and on the chosen point $z_0 \in D$. If, for any $\varepsilon > 0$, there exists a $\delta > 0$ which does the job for all $z_0 \in D$, then f is called *uniformly continuous* on D.

The relation between continuity and convergent sequences is established in the following basic result.

Theorem 2.6.1. *A function f on D is continuous at a point $z_0 \in D$ if and only if, for any sequence $(z_n) \subset D$, $z_n \to z_0$ implies $f(z_n) \to f(z_0)$.*

Function Sequences. Next, we consider sequences (f_n) of functions $f_n : D \to \mathbb{C}$. The sequence (f_n) *converges at* $z \in D$ if the number sequence $(f_n(z))$ converges. If this happens for all points z in D, we say that (f_n) *converges pointwise* on D.

Pointwise convergence is often too weak to get nice results, for example the limit function of a pointwise convergent sequence of continuous functions is not necessarily continuous. Therefore we need a stronger version of convergence.

---

[4] By convention, the "if" in definitions has always the meaning "if and only if".

The sequence (f_n) *converges uniformly* on D to a function $f : D \to \mathbb{C}$ if for every $\varepsilon > 0$ there exists an integer N such that $|f_n(z) - f(z)| < \varepsilon$ for all $n \geq N$ and $z \in D$.

Theorem 2.6.2. *If a sequence of continuous functions f_n converges uniformly on D, then its limit function is continuous on D.*

We remark that the definitions made above also make sense when the domain set D is replaced by a subset A. For functions $f : D \subset \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ the definitions have to be modified accordingly, replacing the Euclidean distance $|z - w|$ with the spherical distance $d(z, w)$.

Number Series. A *series* $\sum_{k=1}^{\infty} a_k$ of complex numbers *converges*, if the sequence (s_n) of its *partial sums* $s_n := \sum_{k=1}^{n} a_k$ converges. The limit $s := \lim_{n\to\infty} s_n$ is said to be the *sum* of the series. The following result is a consequence of the *completeness* of $\mathbb{C}$.

Theorem 2.6.3 (Cauchy Criterion)**.** *A complex series $\sum_{k=1}^{\infty} a_k$ converges if and only if for each $\varepsilon > 0$ there is an N such that $n \geq m \geq N$ implies that*

$$\left| \sum_{k=m}^{n} a_k \right| < \varepsilon.$$

Some manipulations with series (for example rearrangements of the summands) require an even stronger concept of convergence.

A series $\sum_{k=1}^{\infty} a_k$ is said to *converge absolutely*, if $\sum_{k=1}^{\infty} |a_k|$ converges. It follows from Theorem 2.6.3 and the triangle inequality that absolute convergence of a series implies convergence.

The following result on changing the order of summation will be of special importance in later chapters. Here we consider a double sequence a_{jk} of complex numbers with $j, k \in \mathbb{Z}_+$ and form the *iterated sums*

$$\sum_{k=1}^{\infty} \Big(\sum_{j=1}^{\infty} a_{jk} \Big), \qquad \sum_{j=1}^{\infty} \Big(\sum_{k=1}^{\infty} a_{jk} \Big). \tag{2.23}$$

In general, the behavior of both sums can be completely different, but the situation improves if one of the series converges absolutely, i.e., it converges if the a_{jk} are replaced by their absolute values $|a_{jk}|$.

Theorem 2.6.4 (Weierstrass Double Series Theorem)**.** *Let $a_{jk} \in \mathbb{C}$ for $j, k \in \mathbb{Z}_+$. If one of the iterated sums (2.23)* converges absolutely*, then both series converge and*

$$\sum_{k=1}^{\infty} \Big(\sum_{j=1}^{\infty} a_{jk} \Big) = \sum_{j=1}^{\infty} \Big(\sum_{k=1}^{\infty} a_{jk} \Big).$$

Function Series. Pointwise and uniform convergence of *function series* $\sum_{k=1}^{\infty} f_k(z)$ on a set $D \subset \widehat{\mathbb{C}}$ are defined by the corresponding properties of their partial sums. The next theorem is a convenient criterion for proving convergence of such series.

Theorem 2.6.5 (Weierstrass M-Test)**.** *Let* (f_k) *be a sequence of functions defined on* $D \subset \widehat{\mathbb{C}}$*. If there is a sequence of real numbers* M_k *such that* $|f_k(z)| \leq M_k$ *for all* $z \in A$ *and* $k \in \mathbb{Z}_+$*, and* $\sum_{k=1}^{\infty} M_k$ *converges, then* $\sum_{k=1}^{\infty} f_k(z)$ *converges absolutely and uniformly on* D*.*

Landau Notation. A convenient notation for comparing the asymptotic behavior of two functions in a neighborhood of a point $z_0 \in \widehat{\mathbb{C}}$ is provided by the *Landau symbols* O, o, and $\sim$. Following the usual conventions, we write

(i) $f(z) = O\big(g(z)\big)$ if for some C and all z in a neighborhood of z_0

$$|f(z)| \leq C\,|g(z)|,$$

(ii) $f(z) = o\big(g(z)\big)$ if $f(z)/g(z) \to 0$ as $z \to z_0$,

(iii) $f(z) \sim g(z)$ if $f(z)/g(z) \to 1$ as $z \to z_0$.

It goes without saying that the point z_0 where the functions f and g are compared must be specified. The Landau symbols are also used without explicitly mentioning the variable z, i.e., $f = O(g)$, $f = o(g)$, $f \sim g$.

2.7 Some Plane Geometry

The theory of complex functions is a fascinating blend of analysis and geometry. For the convenience of the reader, we assemble some related definitions and results from plane topology in this section. It is not necessary to work through the whole material at once, rather we recommend that the reader consults this section as and when the corresponding concepts become relevant in the ongoing course.

Since the focus of this book is on complex functions, we do not aim to prove *en passant* theorems in plane geometry. Instead we just quote these facts and refer the interested reader to textbooks like Munkres [42] or Henle [25]. For an introduction to geometric concepts in the context of complex analysis we also recommend Chapter II of Palka [52].

Domains. When we shall study analytic functions in the next chapter, it will become clear that their domain sets are necessarily *open* subsets of the complex plane or the Riemann sphere (Lemma 3.3.3). Such sets can be decomposed into even simpler pieces which will be introduced in Definition 2.7.1.

Recall that a set S (in a topological space) is *connected* if whenever S is the union of two non-empty disjoint sets A and B, then (at least) one of these sets contains a point which belongs to the closure of the other.

Definition 2.7.1. A *domain* is a non-void, open and connected subset of the complex plane or the Riemann sphere.

Note that this definition is consistent: any subset of $\mathbb{C}$ which is a domain in $\mathbb{C}$ is also a domain in $\widehat{\mathbb{C}}$. An alternative commonly used term for domain is *“region”*.

The Riemann sphere $\widehat{\mathbb{C}}$, the complex plane $\mathbb{C}$, the unit disk $\mathbb{D}$, the upper half-plane $\mathbb{H} := \{z \in \mathbb{C} : \operatorname{Im} z > 0\}$, and the ring $R := \{z \in \mathbb{C} : 1/2 < |z| < 2\}$ are domains. The next result shows that any open subset of the Riemann sphere is composed of domains.

Proposition 2.7.2. *Any open set in $\widehat{\mathbb{C}}$ is the disjoint union of at most countably many domains.*

The domains which are the building blocks of an open set D according to Proposition 2.7.2 are said to be the *components* of D.

For some purposes even the concept of a domain is too general, and we need a finer classification. For example, a disk and a ring-shaped domain are topologically different. In order to make this distinction precise we need some further preparation.

Paths and Loops. Contrary to a common sense interpretation, a path is a *continuous function* from some closed interval into a topological space. The following definition gives a more detailed description.

Definition 2.7.3. A *path* from a to b in D is a continuous map $\gamma : [\alpha, \beta] \to D$ with $\gamma(\alpha) = a$, $\gamma(\beta) = b$, and $\alpha < \beta$. The points a and b are called the *initial point* and the *terminal point* of γ, respectively. If the initial and the terminal points coincide we speak of a *closed path* or a *loop*. The image set $[\gamma] := \gamma([\alpha, \beta])$ is said to be the *trace* (or *trajectory*) of γ.

The function γ is also referred to as a *parametrization* of the trace $[\gamma]$. Being the continuous image of an interval, the trace of a path is a compact connected set.

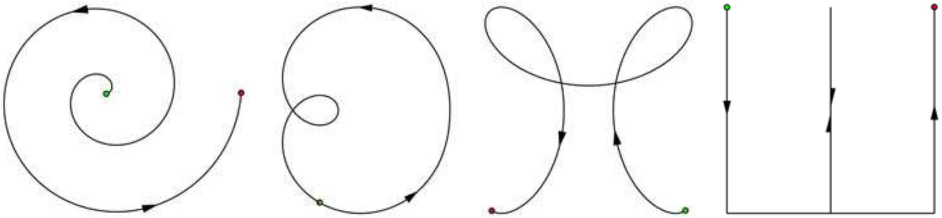


Figure 2.30: Traces of four paths with direction of increasing parameter indicated

A path is an *oriented* object. When we depict the trace of a path, the direction in which the point $\gamma(t)$ traverses $[\gamma]$ is often indicated by an arrow. Figure 2.30

shows a few examples. The green and the red dots are the initial point a and the terminal point b, respectively. These two points are also known simply as the *endpoints* of the path, and we say that γ *joins* (or *connects*) a with b. If γ is a loop, its common initial and terminal point is referred to as the *base point* of γ.

Remark 2.7.4. We do not introduce the concept of a *curve* here, because there is no need for it, at least for the moment. One should be aware that there is no general convention about the meaning of this notion. Some authors consider "curve" and "path" as synonyms, some call the trace of a path a curve, and others define a curve as an equivalence class of paths. In order to avoid confusion, we shall try to avoid using this term, except when we talk about *Jordan curves*, a notion which will be made precise below.

There are some standard paths which will come up often in our discussion. If z_1 and z_2 are two (not necessarily distinct) points in the plane, we denote by $[z_1, z_2]$ the path $\gamma(t) = z_1 + t(z_2 - z_1)$ with $t \in [0, 1]$. So $[z_1, z_2]$ denotes a path with initial point z_1 and terminal point z_2, but we use the same notation also for its (oriented) trace, the *segment* $[z_1, z_2]$.

Speaking of the *standard parametrization of a circle* with center a and radius r we usually mean the path $\gamma(t) = a + r\,\exp(2\pi i t)$ with $t \in [0, 1]$. Sometimes we shall also use $\gamma(t) = a + r\,\exp(it)$ with $t \in [0, 2\pi]$.

Simple Paths. Without additional assumptions, paths may have unexpected properties, which might be inconvenient under certain circumstances. For example, the space filling *"Peano curve"* (which is a path in our terminology) is a continuous surjective mapping of $[0, 1]$ onto the unit square $[0, 1] \times [0, 1]$. It is constructed as a (uniform) limit of a sequence (γ_n) of paths three of which are shown in Figure 2.31.[5] Such weird objects can be excluded by additional assumptions.

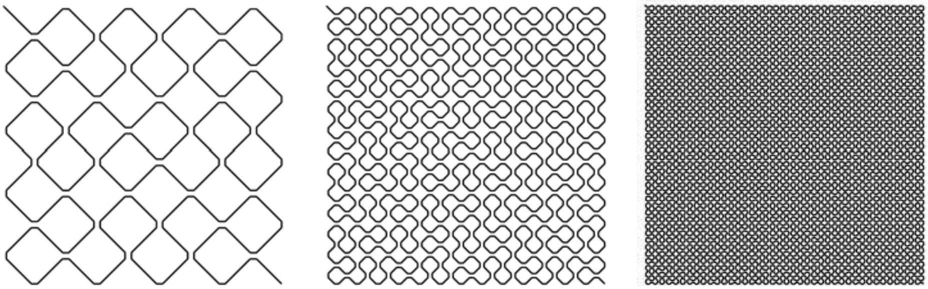


Figure 2.31: Three stages in the construction of the Peano curve

Definition 2.7.5. A path $\gamma : [\alpha, \beta] \to D$ is called *simple*if $\gamma(s) = \gamma(t)$ with $s < t$ implies that $s = \alpha$ and $t = \beta$.

---

[5] The curves depicted are computed using the MATLAB routine `peano.m` by Andreas Klimke.

This somewhat artificial definition allows for two possibilities: either $\gamma(\alpha) \neq \gamma(\beta)$, (in which case γ is not closed) and γ is injective, or $\gamma(\alpha) = \gamma(\beta)$ (i.e., γ is a loop) and the restriction of γ to $[\alpha, \beta)$ is injective.

Change of Parameter. As it turns out, the choice of the parameter interval is not that essential for the definition of a path. If $\gamma_1 : [\alpha_1, \beta_1] \to D$ is a path in D and $\tau : [\alpha_2, \beta_2] \to [\alpha_1, \beta_1]$ is a linear mapping with $\tau(\alpha_2) = \alpha_1$ and $\tau(\beta_2) = \beta_1$, then the path $\gamma_2 := \gamma_1 \circ \tau$ is a path in D with parameter interval $[\alpha_2, \beta_2]$. We call γ_2 a *linear reparametrization* of γ_1.

In fact there are no "essential differences" between γ_1 and γ_2, and none of the results which we shall encounter is affected by linear reparametrization – the sceptical reader is invited to formulate and prove the corresponding results whenever he or she has doubts.

Any path γ admits a (unique) linear reparametrization with the parameter interval $[0, 1]$. In what follows we shall often assume that paths are given in this *normalized form.*

Path Manipulations. Next we introduce some manipulations with paths. The *reversed path* (or *negative* path) γ^- of $\gamma : [\alpha, \beta] \to \mathbb{C}$ is defined on $[\alpha, \beta]$ by $\gamma^-(t) := \gamma(\alpha + \beta - t)$.

If the terminal point of $\gamma_1 : [\alpha_1, \beta_1] \to \mathbb{C}$ coincides with the initial point of $\gamma_2 : [\alpha_2, \beta_2] \to \mathbb{C}$, the *concatenation* (also denoted as *sum*) $\gamma := \gamma_1 \oplus \gamma_2$ of γ_1 and γ_2 is defined on $[\alpha_1, \beta_1 + \beta_2 - \alpha_2]$ by

$$\gamma(t) := \begin{cases} \gamma_1(t) & \text{if} \quad t \in [\alpha_1, \beta_1] \\ \gamma_2(t + \alpha_2 - \beta_1) & \text{if} \quad t \in (\beta_1, \beta_1 + \beta_2 - \alpha_2]. \end{cases}$$

This "sum" is associative, but not commutative. Further we set $\gamma_1 \ominus \gamma_2 := \gamma_1 \oplus \gamma_2^-$, provided that the right-hand side makes sense.

When working with normalized paths γ_1 and γ_2 defined on $[0, 1]$, we use a modified definition of concatenation, where the parameter interval of $\gamma_1 \oplus \gamma_2$ is rescaled to $[0, 1]$ so that the resultant path is also a normalized one.

Further Properties. Sometimes we shall need paths with specific properties. Three such classes are described in the next definition.

Definition 2.7.6. A path γ is said to be *smooth*, if $\operatorname{Re} \gamma$ and $\operatorname{Im} \gamma$ are continuously differentiable functions. A *polygonal path* is the sum $\gamma = \gamma_1 \oplus \ldots \oplus \gamma_n$ of segments $\gamma_k = [z_{k-1}, z_k]$. A polygonal path is called *paraxial* if each of its segments is parallel to the real or the imaginary axis.

Note that smoothness of a path does not necessarily imply that its trace looks like a smooth curve.

Figure 2.32 shows the traces of three (polygonal) paraxial paths. They are members of a sequence (γ_n) which converges (uniformly) to the so-called *Hilbert curve*, another continuous surjective mapping of an interval onto a square.[6]

---

[6] The depicted curves are computed using the MATLAB routine `hilbert.m` by Federico Forte.

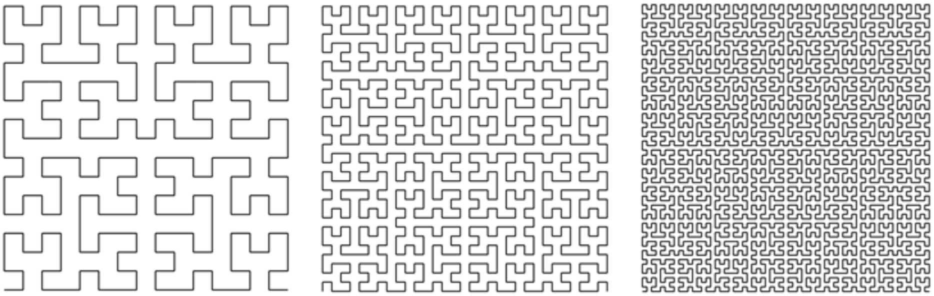


Figure 2.32: Three paraxial paths (stages in the construction of the Hilbert curve)

Path Covering. The following technical lemma will be useful in several proofs below. More importantly, it will be crucial for our discussion on analytic continuation of complex functions in Section 3.6. Roughly speaking the result says that the trace of a path in a domain D can be covered by a finite collection of well-overlapping disks in D as is illustrated in Figure 2.33. For later applications we need some additional requirements on the location of the centers and the covering properties of the disks; they are made precise in the following definition.

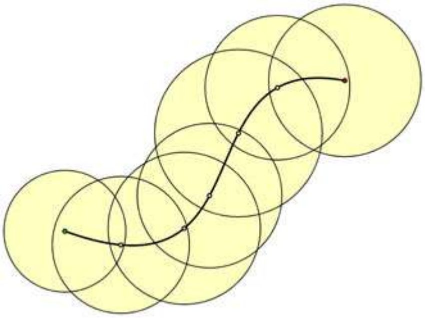


Figure 2.33: Covering a path by disks

Definition 2.7.7. Let $\gamma : [0,1] \to \mathbb{C}$ be a path in the complex plane. A *chain of disks covering* γ is a finite sequence $(D_0, D_1, \ldots, D_n)$ of open disks D_k with the following properties:

(i) There exists a partition $0 = t_0 < t_1 < \ldots < t_n = 1$ of the interval $[0,1]$ such that $\gamma(t_k)$ is the center of D_k for $k = 0, 1, \ldots, n$.

(ii) The section of γ between $\gamma(t_{k-1})$ and $\gamma(t_{k+1})$ is contained in D_k, more precisely,

$$\begin{aligned} &\gamma(t) \subset D_0, \quad t_0 \le t \le t_1, \\ &\gamma(t) \subset D_k, \quad t_{k-1} \le t \le t_{k+1}, \qquad (k = 1, \ldots, n-1), \\ &\gamma(t) \subset D_n, \quad t_{n-1} \le t \le t_n. \end{aligned}$$

Lemma 2.7.8 (Path Covering Lemma)**.** *Let γ be a path in the domain $D \subset \mathbb{C}$. Then there exists a chain of disks which is contained in D and covers γ. Moreover, the radii of all disks can be chosen to be of the same size and arbitrarily small.*

Proof. Since γ is continuous on the compact interval $[0,1]$, its trace $[\gamma]$ is a compact subset of D. The complement of D in $\mathbb{C}$ is closed, and hence the distance d between $[\gamma]$ and $\mathbb{C} \setminus D$ is positive. If $0 < r < d$, then all disks with radius r and centers on $[\gamma]$ are contained in D. Because γ is uniformly continuous, there exists a positive number δ such that $s, t \in [0,1]$ and $|s - t| < \delta$ imply that $|\gamma(s) - \gamma(t)| < r$. So all requirements are satisfied if the partition $0 = t_0 < t_1 < \ldots < t_n = 1$ is chosen such that $t_k - t_{k-1} < \delta$. $\square$

Homotopy. Next we sketch a powerful technique, which formalizes the intuitive concept of a *continuous deformation* of a path γ_0 into a path γ_1. The idea is to embed γ_0 and γ_1 into a family of paths $\gamma_s : t \mapsto h(s,t)$, which depend continuously on the parameter s. The technical term for this procedure is *homotopy.* We shall consider two situations: deformations of paths with fixed endpoints and, somewhat later, free deformations of *closed* paths (loops) without this restriction. To simplify notation we assume that all paths are defined on $[0,1]$, which can always be achieved by a linear reparametrization.

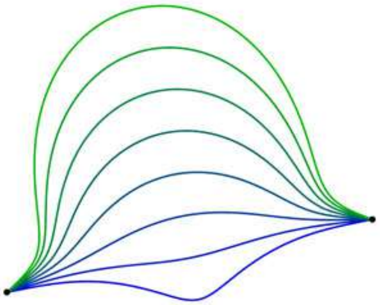


Figure 2.34: Homotopic paths

Definition 2.7.9. Two paths $\gamma_0 : [0,1] \to D$ and $\gamma_1 : [0,1] \to D$ from a to b are said to be *homotopic in D with fixed endpoints* if there exists a continuous function $h : [0,1] \times [0,1] \to D$ which satisfies the following conditions (i)-(iv):

$$\begin{array}{lll} \text{(i)} & h(0,t) = \gamma_0(t), & 0 \le t \le 1 \\ \text{(ii)} & h(1,t) = \gamma_1(t), & 0 \le t \le 1 \\ \text{(iii)} & h(s,0) = a, & 0 \le s \le 1 \\ \text{(iv)} & h(s,1) = b, & 0 \le s \le 1. \end{array}$$

The function h will be referred to as a *homotopy* from γ_0 to γ_1.

If there is no danger of confusion we shall just speak of homotopic paths without mentioning that the endpoints are fixed.

Figure 2.34 illustrates that the family (γ_s) of intermediate paths $\gamma_s := h(s,.)$ with $s \in [0,1]$ is a continuous deformation of γ_0 into γ_1. The traces of γ_s are colored according to the homotopy parameter s.

Note that the coincidence of the traces $[\gamma_0]$ and $[\gamma_1]$ does *not* guarantee that γ_0 and γ_1 are homotopic. For example the paths γ_0 and γ_1 in the domain $\mathbb{C} \setminus \{0\}$, defined by

$$\gamma_0(t) := \mathrm{e}^{2\pi \mathrm{i} t}, \quad \gamma_1(t) := \mathrm{e}^{4\pi \mathrm{i} t}, \qquad t \in [0,1],$$

have the same trace (the unit circle) though they are not homotopic in $\mathbb{C}\setminus\{0\}$.

It is not difficult to see that "being homotopic" is an equivalence relation. So the set of all paths in D with fixed endpoints a and b splits into classes of homotopic paths. Note also that reparametrization of a path does not change its homotopy class, no matter what the domain D is.

Definition 2.7.10. A path $\gamma_1 : [0,1] \to D$ is said to be a *reparametrization* of the path $\gamma_0 : [0,1] \to D$ if there exists a continuous function $\varphi : [0,1] \to [0,1]$ with $\varphi(0) = 0$ and $\varphi(1) = 1$ such that $\gamma_1 = \gamma_0 \circ \varphi$.

Lemma 2.7.11. *Let γ be a path in D. Then any reparametrization $\gamma\circ\varphi$ is homotopic to γ in D.*

Proof. The function h defined by $h(s,t) := \gamma\big((1-s)\,t + s\,\varphi(t)\big)$ maps $[0,1]\times[0,1]$ into D and is a homotopy from γ to $\gamma\circ\varphi$. □

Homotopic Paths with Specific Properties. Technically it is of great importance that any path in D can be approximated by homotopic paths with specific properties. In particular, the next lemma shows that any class of homotopic paths contains a smooth and a paraxial path (see Definition 2.7.6).

Lemma 2.7.12. *Let $\gamma : [0,1] \to D$ be a path in an open set $D \subset \mathbb{C}$. Then there exist a smooth path $\widetilde{\gamma} : [0,1] \to D$ and a paraxial path $\widehat{\gamma} : [0,1] \to D$ which are homotopic to γ in D. For each positive ε both paths can be chosen such that*

$$\max_{t\in[0,1]} |\gamma(t) - \widetilde{\gamma}(t)| < \varepsilon, \qquad \max_{t\in[0,1]} |\gamma(t) - \widehat{\gamma}(t)| < \varepsilon.$$

Proof. By the path covering lemma (Lemma 2.7.8, see page 49 for the notation used in the following), γ can be covered by a sequence of disks D_k with radii less than $\varepsilon/2$. Let $0 = t_0 < t_1 < \ldots < t_n = 1$ be the subdivision of the parameter interval $[0,1]$, and denote by $z_k = \gamma(t_k)$ the centers of the covering disks D_k. Then the restriction γ_k of γ to $[t_{k-1}, t_k]$ is homotopic in D_k (and hence in D) to the line segment $[z_{k-1}, z_k]$ for all $k = 1,\ldots,n$. This induces a homotopy between γ and the polygonal path $\widehat{\gamma} := [z_0, z_1] \oplus \ldots \oplus [z_{n-1}, z_n]$. Smoothing the function $\widehat{\gamma}$ at the points t_k appropriately, we obtain a smooth path $\widetilde{\gamma}$ which is homotopic to $\widehat{\gamma}$ and hence to γ. Finally, the segments $[z_{k-1}, z_k]$ are homotopic in D_k to the sum $[z_{k-1}, \operatorname{Re} z_k + \mathrm{i}\operatorname{Im} z_{k-1}] \oplus [\operatorname{Re} z_k + \mathrm{i}\operatorname{Im} z_{k-1}, z_k]$ of two segments which are parallel to the real and imaginary axis, respectively. □

Connectedness Revisited. The following result is sometimes used as a definition of connectedness for open sets. Note that the "only if" conclusion fails when D is not open.

Proposition 2.7.13. *An open set $D \subset \widehat{\mathbb{C}}$ is connected if and only if any two points a and b in D can be joined by a path γ in D.*

In conjunction with Lemma 2.7.12 we conclude that γ can be chosen to be smooth or paraxial.

Simply Connected Domains. The definition of homotopy implies that the initial points and the terminal points of two homotopic paths must coincide. In general the converse it not true, but the next definition singles out a class of domains for which the endpoint condition is also sufficient for the homotopy of paths.

Definition 2.7.14. A domain D is called *simply connected* if any two paths $\gamma_0 \subset D$ and $\gamma_1 \subset D$ with $\gamma_0(0) = \gamma_1(0)$ and $\gamma_0(1) = \gamma_1(1)$ are homotopic in D. Domains which are not simply connected are said to be *multiply connected*.

The Riemann sphere $\widehat{\mathbb{C}}$, the complex plane $\mathbb{C}$, and the unit disk $\mathbb{D}$ are simply connected, while the *punctured unit disk*

$$\dot{\mathbb{D}} := \{z \in \mathbb{C} : 0 < |z| < 1\},$$

and the ring domain R

$$R := \{z \in \mathbb{C} : 1/2 < |z| < 2\}$$

are not simply connected. The two paths γ_0 and γ_1 depicted in Figure 2.35 have the same initial and the same terminal point, but they are not homotopic in R.

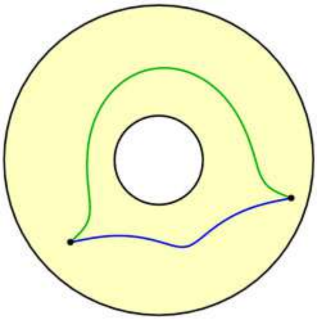


Figure 2.35: Non-homotopic paths

Homotopy of Loops. For closed paths there is a second notion of homotopy which is more general because it does not require that the endpoints be kept fixed.

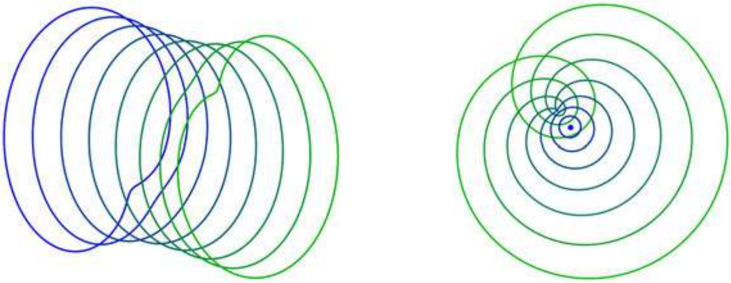


Figure 2.36: Traces of two families of freely homotopic closed paths

Definition 2.7.15. Two loops γ_0 and γ_1 in a set D are called *freely homotopic in D* if there exists a continuous function $h : [0,1] \times [0,1] \to D$ such that

(i) $h(0,t) = \gamma_0(t), \quad 0 \le t \le 1$

(ii) $h(1,t) = \gamma_1(t), \quad 0 \le t \le 1$

(iii) $h(s,0) = h(s,1), \quad 0 \le s \le 1.$

A loop which is freely homotopic in D to a constant path is said to be *null-homotopic* in D.

Figure 2.36 illustrates the definition. Both pictures visualize a freely homotopic family of closed path. The homotopy depicted on the right-hand side contracts the green curve to the blue point.

Lemma 2.7.16. *For any path γ in D the loop $\gamma \oplus \gamma^-$ is null-homotopic in D to its base point.*

Proof. We contract $\gamma \oplus \gamma^-$ along $[\gamma]$ to its base point $\gamma(0)$ as illustrated on the left of Figure 2.37. More formally, the function h defined for $s, t \in [0,1]$ by

$$h(s,t) = \begin{cases} \gamma(2st) & \text{if} \quad 0 \le t \le 1/2 \\ \gamma\big(2s(1-t)\big) & \text{if} \quad 1/2 < t \le 1 \end{cases}$$

is continuous on $[0,1] \times [0,1]$, its range is contained in $[\gamma]$, and hence in D, and it satisfies $h(0,.) = \gamma(0)$ and $h(1,.) = \gamma \oplus \gamma^-$. □

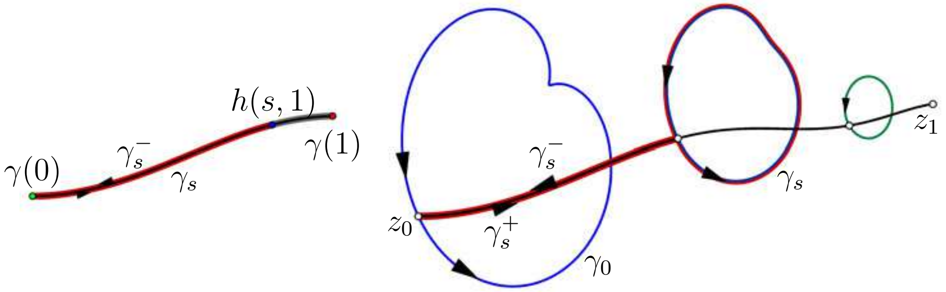


Figure 2.37: Illustrations to the proofs of Lemma 2.7.16 and Lemma 2.7.17

The following result shows that a null-homotopic loop can always be contracted to its base point.

Lemma 2.7.17. *If a loop with base point z_0 is null-homotopic in D, then it is also homotopic with fixed endpoints to the constant path z_0.*

Proof. Let h be a homotopy from the given path γ_0 to a point z_1. We define γ_s and γ_s^+ by

$$\gamma_s(t) := h(s,t), \quad \gamma_s^+(t) = h(st,0), \qquad s,t \in [0,1],$$

and set $\gamma_s^- := (\gamma_s^+)^-$. Then the path γ_s^+ lies in D and connects z_0 with the moving base point $z_s := h(s,0) = h(s,1)$ of the loop γ_s, as shown in Figure 2.37 (right). The family of loops

$$\gamma_s^* := \gamma_s^+ \oplus \gamma_s \oplus \gamma_s^-$$

has fixed base point z_0 and all paths in this family are homotopic in D (Figure 2.37 shows the trace of one such path γ_s^* in red). Now γ_0 is homotopic to γ_0^*, γ_0^* is homotopic to γ_1^*, and γ_1^* is homotopic to $\gamma_1^+ \oplus \gamma_1^-$, and by Lemma 2.7.16 the latter is homotopic to the base point z_0. ☐

Lemma 2.7.18. *A domain is simply connected if and only if any loop in D is null-homotopic in D.*

Proof. 1. Assume that D is simply connected and $\gamma : [0,1] \to D$ is a loop. If γ_1 and γ_2 denote the restrictions of γ to the intervals $[0,1/2]$ and $[1/2,1]$, respectively, then $\gamma = \gamma_1 \oplus \gamma_2$. The paths γ_2 and γ_1^- have the same initial and the same terminal points and hence they are homotopic in D with fixed endpoints. This implies that γ and $\gamma_1 \oplus \gamma_1^-$ are homotopic in D (with fixed endpoints and also freely). Now apply Lemma 2.7.16.

2. Let D be a domain and suppose that any loop in D is null-homotopic. If γ_0 and γ_1 are paths with initial point a and terminal point b, then $\gamma := \gamma_0 \oplus \gamma_1^-$ is a loop with base point a and $\gamma_0 = \gamma \oplus \gamma_1$. By assumption and Lemma 2.7.16 γ is homotopic with fixed endpoints to the constant path a via a family of paths γ^s. This induces the homotopic family of paths $\gamma_s := \gamma^s \oplus \gamma_1$ from γ_0 to γ_1. ☐

Winding Numbers. We now introduce a geometric characteristic of loops which describes how many times they "wind around" some point in the plane. Among other applications, this concept will be essential for understanding the complex logarithm function. We begin with an auxiliary result.

Lemma 2.7.19. *Let $\gamma : [0,1] \to \mathbb{C} \setminus \{0\}$ be a path. Then there exist continuous functions $a : [0,1] \to \mathbb{R}$ and $r : [0,1] \to \mathbb{R}_+$, such that*

$$\gamma(t) = r(t)\, \mathrm{e}^{\mathrm{i}\, a(t)}. \tag{2.24}$$

Proof. The function $r(t) := |\gamma(t)|$ is continuous and positive. So the proof amounts to finding an appropriate argument $a(t)$ of $\gamma(t)$ such that $t \mapsto a(t)$ is continuous. For this purpose we use the path covering lemma with $D := \mathbb{C} \setminus \{0\}$ (see page 49 for notation).

At the initial point of γ we choose the principal branch of the argument, $a(0) := \operatorname{Arg} \gamma(0)$. If $t \in [t_0, t_1]$, all points $\gamma(t)$ lie in the disk D_0. Since D_0 does not contain the origin, it is contained in a sector with vertex at 0 and opening angle less than π. Consequently the argument $a(t) = \arg \gamma(t)$ can be chosen such that $|a(t) - a(0)| < \pi/2$, which yields a continuous function a on $[0, t_1]$.

Suppose that such a function has already been constructed on some interval $[0, t_k]$. Then it can be prolongated to $[0, t_{k+1}]$ by choosing $a(t) = \arg \gamma(t)$ on $[t_k, t_{k+1}]$ such that $|a(t) - a(t_k)| < \pi/2$, which is possible since $\gamma(t) \in D_k$ and $0 \notin D_k$. By induction, a can be extended to all of $[0,1]$. ☐

Any continuous function a satisfying (2.24) is called a *continuous branch* of the argument along the path γ. The difference of two such functions a_1 and a_2 on $[0,1]$ is a constant integral multiple of 2π.

If a is a continuous branch of the argument along a loop, then $a(1) - a(0)$ is an integral multiple of 2π which does not depend on the special choice of the branch a.

Definition 2.7.20. Let γ be a loop in $\mathbb{C} \setminus \{0\}$ and denote by a a continuous branch of the argument along γ. Then the integer

$$\operatorname{wind} \gamma := \frac{1}{2\pi} \big(a(1) - a(0) \big)$$

is called the *winding number* (or *index*) of γ. If $z_0 \in \mathbb{C}$ and γ is a loop in $\mathbb{C} \setminus \{z_0\}$, the *winding number of* γ *about* z_0 is defined by

$$\operatorname{wind}(\gamma, z_0) := \operatorname{wind}(\gamma - z_0).$$

Figure 2.38 shows the trace of a loop with winding number 1 about the origin (marked by a black dot). The coloring visualizes the winding number of γ about points z_0 in the different components of $\mathbb{C} \setminus [\gamma]$.

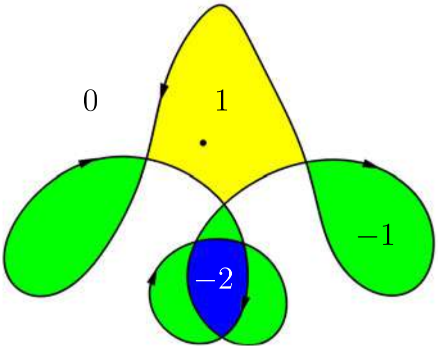


Figure 2.38: Winding numbers $\operatorname{wind}(\gamma, z_0)$

Stability of Winding Numbers. Next we prove the intuitive fact that small perturbations of a loop do not change its winding number. In order to obtain a quantitative version of this result, we measure the distance between two paths γ_0 and γ defined on $[0,1]$ by

$$\|\gamma - \gamma_0\| := \max_{t \in [0,1]} |\gamma(t) - \gamma_0(t)|.$$

Lemma 2.7.21. *Let* $\gamma_0 : [0,1] \to\in \mathbb{C} \setminus \{0\}$ *be a loop, and denote by* d *the distance of its trace* $[\gamma_0]$ *from the origin. Then for all loops* $\gamma : [0,1] \to \mathbb{C}$ *with* $\|\gamma - \gamma_0\| < d$

$$\operatorname{wind} \gamma = \operatorname{wind} \gamma_0.$$

Proof. Since $[\gamma_0]$ is a compact subset of $\mathbb{C} \setminus \{0\}$, its distance d from the origin is positive. Then $|\gamma(t) - \gamma_0(t)| < d$ implies that $\gamma(t)/\gamma_0(t)$ lies in the right half-plane.

Let a_0 be a continuous branch of the argument along γ_0. If we chose a continuous branch of the argument along γ such that $|a(0) - a_0(0)| < \pi/2$, then $|a(t) - a_0(t)| < \pi/2$ for all $t \in [0,1]$. Invoking the triangle inequality we see that $\big| \big(a(1) - a(0)\big) - \big(a_0(1) - a_0(0)\big) \big| < \pi$, and since this number is an integral multiple of 2π it must be zero. □

Homotopic Loops in Punctured Domains. The next result describes a situation where the winding number alone allows one to decide if two loops are homotopic.

Lemma 2.7.22. *Let $D \subset \mathbb{C}$ be a simply connected domain and $z_0 \in D$. Then two loops γ_0 and γ_1 are homotopic in the punctured domain $D \setminus \{z_0\}$ if and only if they have the same winding number about z_0.*

Proof. That homotopic loops have the same winding number follows easily from Lemma 2.7.21 and a compactness argument. Though the converse is plausible, its proof is non-trivial and we refer to the literature (for instance Hatcher [24]). ☐

Jordan Arcs and Jordan Curves. Next we turn our attention to geometric properties of traces of simple paths. This is one of the rare occasion where we resort to the notion of a curve.

Definition 2.7.23. A *Jordan curve* is the trace of a simple loop, the trace of a simple non-closed path is called a *Jordan arc*.

In the definition above we have avoided specifying the target set of the loop γ. Usually we assume that γ maps into $\mathbb{C}$, but occasionally we also admit more generally that $\gamma : [\alpha, \beta] \to \widehat{\mathbb{C}}$, in which case we speak of a Jordan curve on the Riemann sphere. The minor modifications which are needed in the formulation of the corresponding results are left to the reader.

As geometric objects, Jordan curves are not that tame as one might expect from their definition. Indeed their menagerie hosts such badly behaved entities like the fractal *Koch curve* (the boundary of the Koch snowflake) and the *Osgood curve* which has a positive area (see Sagan [60]). So it is not surprising that the following celebrated theorem is not easy to prove.

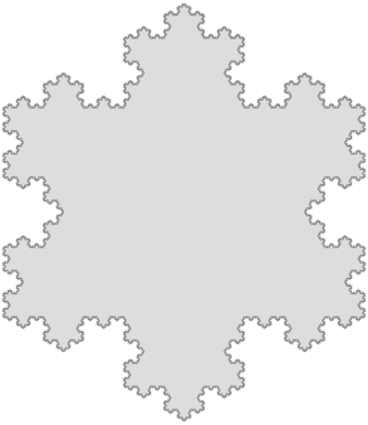


Figure 2.39: The Koch snowflake

Proposition 2.7.24 (Jordan Curve Theorem)**.** *Let J be a Jordan curve in $\mathbb{C}$. Then its complement $\mathbb{C} \setminus J$ is the union of two domains, one of which is bounded and simply connected (the* interior *of J), and the other one is unbounded (the* exterior *of J). The Jordan curve J is the common boundary of its interior and its exterior.*

A relatively short and elementary proof is given by Maehara [39], another instructive proof can be found in Henle [25]. For an account of the history of the theorem we refer to Hales [23].

We denote the interior and the exterior of J by $\operatorname{int} J$ and $\operatorname{ext} J$, respectively. If γ is a simple loop which parameterizes J, then its winding number $\operatorname{wind}(\gamma, z_0)$ about z_0 is well defined for every point $z_0 \in \mathbb{C} \setminus J$ and we have

$$\operatorname{wind}(\gamma, z_0) = \pm 1 \quad \text{if} \quad z_0 \in \operatorname{int} J, \qquad \operatorname{wind}(\gamma, z_0) = 0 \quad \text{if} \quad z_0 \in \operatorname{ext} J, \tag{2.25}$$

with one and the same sign in the first equality for all $z_0 \in \operatorname{int} J$.

Orientation. We say that γ is *positively oriented*, if $\operatorname{wind}(\gamma, z_0) = 1$ for (one and then for all) $z_0 \in \operatorname{int} J$. In this case we shall also call the Jordan curve J parameterized by γ *positively oriented.*

Example 2.7.1 (Triangles). Let z_1, z_2, z_3 be three (not necessarily distinct) points in the complex plane. The *triangle* $\triangle := \triangle(z_1, z_2, z_3)$ is the (closed) convex hull of z_1, z_2, z_3, and

$$\gamma := [z_1, z_2] \oplus [z_2, z_3] \oplus [z_3, z_1]$$

is a path along its boundary $\partial\triangle$. If the interior $\operatorname{int}\triangle$ of $\triangle$ is not empty, we pick a point $z_0 \in \operatorname{int}\triangle$ and define the *standard parametrization* of $\partial\triangle$ as

$$\delta := \begin{cases} \gamma & \text{if} \quad \operatorname{wind}(\gamma, z_0) > 0 \\ \gamma^- & \text{if} \quad \operatorname{wind}(\gamma, z_0) < 0. \end{cases}$$

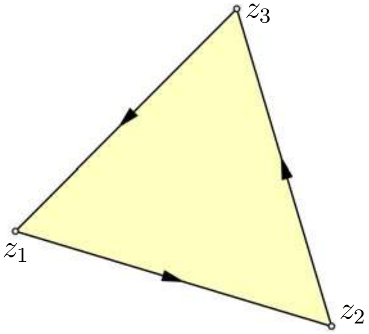


Figure 2.40: A Triangle $\triangle(z_1, z_2, z_3)$

Note that δ is a positively oriented parametrization of $\partial\triangle$.

Jordan Domains. A *Jordan domain* is the interior of a Jordan curve. It is clear that the closure of a Jordan domain is a compact set, but in fact it has a much nicer property.

Proposition 2.7.25. *The closure of a Jordan domain is homeomorphic to the closed unit disk.*

Spelled out explicitly, this means that there is a continuous bijective mapping φ of $\overline{\mathbb{D}}$ onto $K := \operatorname{int} J \cup J$ with continuous inverse. Since such a mapping sends interior points to interior points, and boundary points to boundary points, the images $\varphi(\mathbb{D})$ and $\varphi(\mathbb{T})$ of the unit disk and the unit circle are the Jordan domain $\operatorname{int} J$ and the Jordan curve J, respectively. In fact Jordan curves are precisely the homeomorphic images of circles.

Contractability. The last result in this section tells us that every Jordan curve is "contractible" to any point in its interior. The precise meaning of this statement is made clear in the following result.

Lemma 2.7.26. *Let γ be a parametrization of a Jordan curve J and let $z_0 \in \operatorname{int} J$. Then γ is null-homotopic in $K := \operatorname{int} J \cup J$ to the constant path z_0.*

Proof. Let $\varphi : \overline{\mathbb{D}} \to K$ be a homeomorphism between the closed unit disk and K according to Proposition 2.7.25. Then $\gamma_0 := \varphi^{-1} \circ \gamma$ is a loop which parameterizes the unit circle $\mathbb{T}$. The homotopy

$$h(s,t) := (1-s)\,\gamma_0(t) + s\,w_0, \qquad s,t \in [0,1]$$

contracts γ_0 in the (convex) closed unit disk $\overline{\mathbb{D}}$ to the point $w_0 := \varphi^{-1}(z_0)$. Since φ is a homeomorphism, $\varphi \circ h$ is a homotopy of γ to the point z_0 in K. $\square$

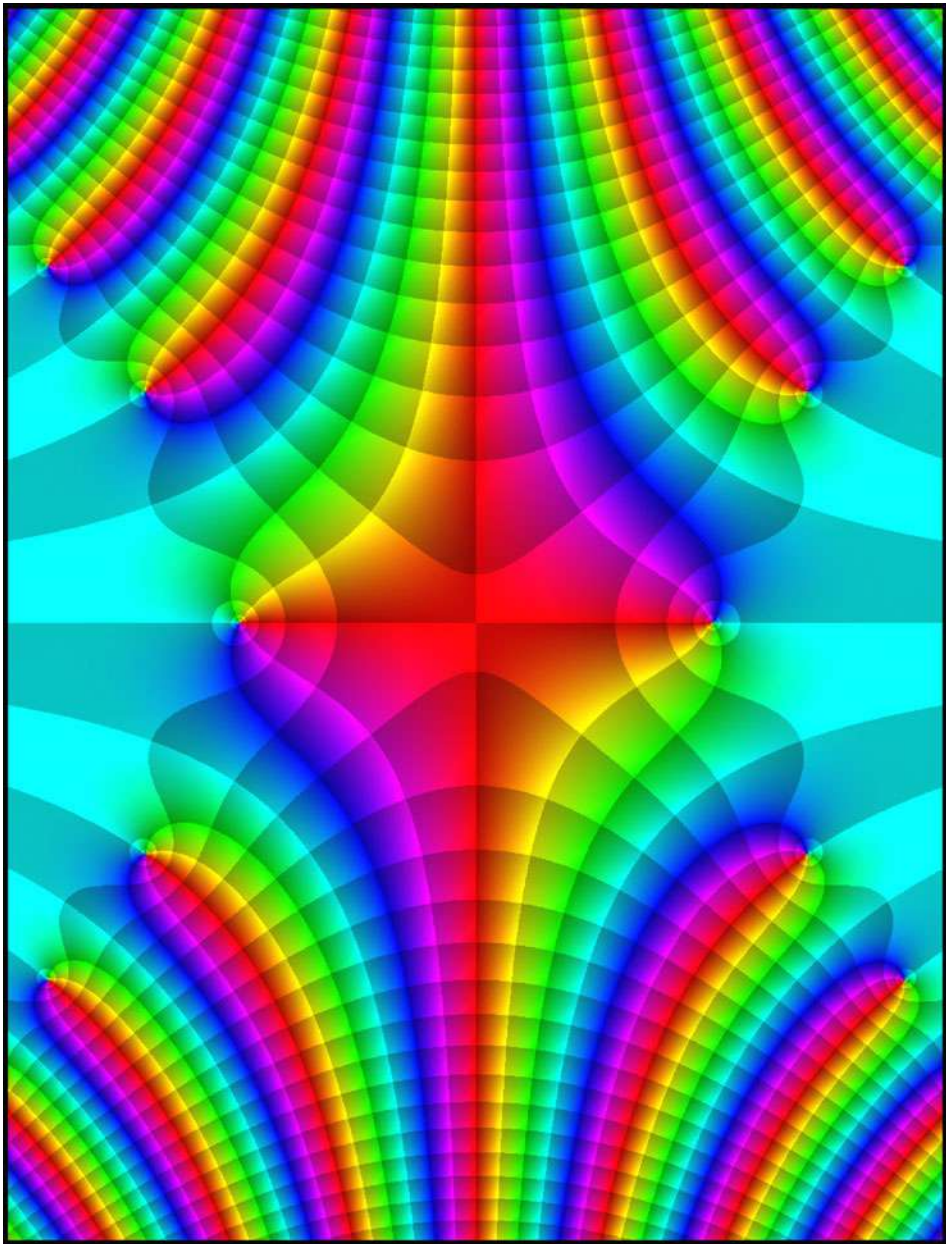


Figure 3.1: Enhanced phase portrait of the function $f(z) = \mathrm{e}^{-z^2} - 1$

Chapter 3

Analytic Functions

In this chapter we shall get introduced to the most important entities of this book: analytic functions. Though they form just a small minority within the class of all complex functions, they are of crucial importance both in theory and in applications. Moreover, they possess a number of fascinating and somewhat unexpected properties which we shall explore in this chapter.

Unlike most contemporary textbooks, where the concept of complex differentiability is used as the entry point to analytic functions, the material in this chapter is developed without recourse to differentiation and integration. Rather, we follow another route with a long historical tradition, namely Karl Weierstrass' approach via power series.

Albeit somewhat technical, working with power series also has some advantages. For example, sums of power series are automatically infinitely differentiable, so that we do not need to worry about additional regularity assumptions, like continuity of the derivative. Moreover, power series provide us with a rich stock of interesting functions right from the beginning.

Emphasizing on the constructive aspects, we try to explore the world of functions step by step: starting with simple functions (powers, polynomials, rational functions), going on to power series, using analytic continuation to obtain functions on larger domains, and finally ending up with functions on Riemann surfaces.

This chapter is organized as follows: after investigating polynomials and rational functions in Section 3.1, we develop the basic results of power series in Section 3.2. Readers who are distracted by the technical details may just read the main results and skip the more involved computational parts. In any case, this text is not the place for an in-depth treatment of the topic. We recommend Henrici's book [26] to readers who wish to gain a deeper understanding of the rich subject of power series.

In Section 3.3 we introduce analytic functions as those complex functions which can be locally represented by power series. A central result is the Weierstrass rearrangement theorem, which will play a basic role in Section 3.6. Furthermore,

E. Wegert, *Visual Complex Functions: An Introduction with Phase Portraits*,
DOI 10.1007/978-3-0348-0180-5_3, © Springer Basel 2012

we shall establish local normal forms of analytic functions, and investigate how these are reflected in the phase portraits.

In Section 3.4 we obtain a number of results about analytic functions, viz. the uniqueness principle, the maximum modulus principle, the open mapping principle, and the argument principle. Several further properties will be derived from these basic facts.

In Section 3.5 we extend the concept of analyticity so as to allow the point at infinity to belong to the domain or the range of a function. Readers who are not willing to spend their time with this more general class of analytic functions may skip Section 3.5 at the first reading.

The main theme of Section 3.6 is the Weierstrass disk chain method, a constructive approach for extending an analytic function to a larger domain. Using this technique we shall explore analytic functions along paths and gain new insights into their global structure.

During this process, Riemann surfaces arise naturally and (almost) unavoidably, but including this rich subject would go beyond the scope of this chapter. Having finished Section 3.6, the interested reader may go directly to Chapter 7 where we shall continue this line of investigation.

3.1 Polynomials and Rational Functions

In this section we study *rational functions*, which can be generated from the constant functions and the identity function $f(z) = z$ solely using the four basic arithmetic operations addition, subtraction, multiplication, and division. All these functions will be defined on the complex plane and then extended to the Riemann sphere.

Power Functions. Besides the Möbius transformations which were introduced in Section 2.3, the simplest rational functions are the *power functions* $f(z) = z^n$ with integer exponent n. These functions are naturally defined for $n \in \mathbb{Z}$ and $z \in \mathbb{C} \setminus \{0\}$ and extended to $\widehat{\mathbb{C}}$ by

$$z^n := \begin{cases} 0 & \text{if} \quad (n > 0 \text{ and } z = 0) \quad \text{or} \quad (n < 0 \text{ and } z = \infty) \\ 1 & \text{if} \quad (n = 0 \text{ and } z = 0) \quad \text{or} \quad (n = 0 \text{ and } z = \infty) \\ \infty & \text{if} \quad (n < 0 \text{ and } z = 0) \quad \text{or} \quad (n > 0 \text{ and } z = \infty). \end{cases}$$

So, if $n > 0$, then z^n has a *zero* at $z = 0$ and a *pole* at $z = \infty$, while for $n < 0$ it has a pole at $z = 0$ and a zero at $z = \infty$. In all these cases, the number $|n|$ is said to be the *order* or *multiplicity* of the zero or the pole, respectively.

Note that we have set $z^0 := 1$ for all $z \in \widehat{\mathbb{C}}$, including $z = 0$ and $z = \infty$. These definitions guarantee that all power functions $f(z) = z^n$ are continuous as mappings $f : \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$. For $n \neq 0$, all points $w \in \widehat{\mathbb{C}}$ but $w = 0$ and $w = \infty$ have exactly $|n|$ pre-images. In a certain sense, the mapping $z \mapsto w = z^n$ wraps the z sphere $|n|$ times around the w sphere.

Some phase portraits of the power functions $f(z) = z^n$ in the *unit square* $\mathbb{U} := \{z \in \mathbb{C} : |\operatorname{Re} z| < 1,\ |\operatorname{Im} z| < 1\}$ are shown in Figure 3.2.

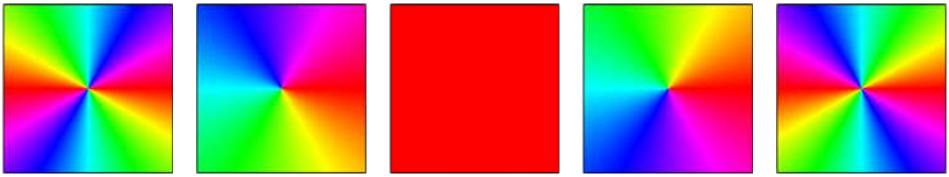


Figure 3.2: Phase portraits of the functions $z^{-2}, z^{-1}, z^0, z^1, z^2$ near $z = 0$

To interpret these pictures we invoke de Moivre's formula. It tells us that the phase of z^n is constant along the rays emerging from the origin. Moreover, when z moves around the unit circle, say in a counterclockwise direction, then the phase $\psi(z^n) = z^n$ rotates on the unit (color) circle with n-times the speed of z. If $n < 0$, then $\psi(z^n)$ travels in the reverse (clockwise) direction of z. Consequently, poles and zeros can be distinguished by the orientation of the colors which appear in their neighborhood: for a zero the colors have the same orientation as on the color circle, for a pole the orientation is reversed. The order of a zero or a pole can also be easily seen, it is just the number of isochromatic rays of one (arbitrarily chosen) color which meet at that point.

Though the reasoning must be slightly modified when we consider poles and zeros at infinity, the above assertions remain valid *in toto*. Figure 3.2 depicts the corresponding phase portraits of z^{-2}, z^{-1}, z^0, z^1, and z^2 in a neighborhood of the point at infinity (see Section 2.5 for details of the construction and the orientation of the pictures).

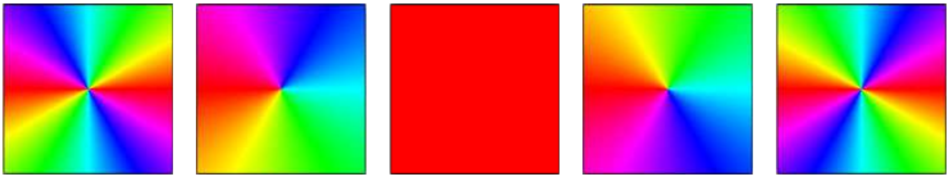


Figure 3.3: Phase portraits of the functions $z^{-2}, z^{-1}, z^0, z^1, z^2$ near $z = \infty$

Polynomials. Power functions are the elementary building blocks of *polynomials*, the most general functions that can be formed using only the operations of addition, subtraction and multiplication. Customarily, polynomials are written in the *normal form*

$$f(z) = a_0 + a_1\, z + a_2\, z^2 + \ldots + a_n\, z^n, \qquad z \in \mathbb{C}. \tag{3.1}$$

The complex numbers $a_0, a_1, \ldots, a_n$ are the *coefficients* of the polynomial. If the leading coefficient a_n does not vanish, which we usually assume, then the number

n is referred to as the *degree* of the polynomial, $n = \deg f$. For the *zero polynomial* $f \equiv 0$ this normalization is impossible, and we set its degree to $-\infty$.

Definition (3.1) is inappropriate for assigning a value to f at infinity, since it would involve undefined expressions like $\infty + \infty$ and, possibly, $0 \cdot \infty$. However, there is still a natural definition of $f(\infty)$. If we wish to extend f to a *continuous* mapping $f : \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$, there can be at most one value for $f(\infty)$ which will do the job. For the moment let us assign

$$f(\infty) := \begin{cases} \infty & \text{if } \deg f \geq 1 \\ a_0 & \text{otherwise.} \end{cases} \tag{3.2}$$

That this is indeed the right choice will be shown in Theorem 3.1.1.

Theorem 3.1.1. *The polynomials defined by* (3.1) *and* (3.2) *are continuous functions* $f : \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$.

Proof. If f is constant, the result is obvious, so assume that $\deg f \geq 1$. The continuity of f at points $z_0 \in \mathbb{C}$ follows from the continuity of $g(z) = z$ and the fact that sum and product of continuous functions are continuous functions. It remains to show that f is continuous at the point at infinity. Since then distances must be measured in the spherical metric, continuity of f at infinity can be rephrased as follows: for any positive ε there exists a positive δ such that $|z| > 1/\delta$ implies $|f(z)| > 1/\varepsilon$. In order to verify this claim, we rewrite $f(z)$ as

$$f(z) = z^n \left(a_n + \frac{a_{n-1}}{z} + \ldots + \frac{a_1}{z^{n-1}} + \frac{a_0}{z^n} \right), \quad z \neq 0. \tag{3.3}$$

Since $|a_n|$ is positive, we find a positive δ such that for all z with $|z| > 1/\delta$

$$\left| a_n + \frac{a_{n-1}}{z} + \ldots + \frac{a_1}{z^{n-1}} + \frac{a_0}{z^n} \right| \geq |a_n| - \left| \frac{a_{n-1}}{z} + \ldots + \frac{a_1}{z^{n-1}} + \frac{a_0}{z^n} \right| \geq \frac{|a_n|}{2}.$$

Choosing δ small enough, so that $\delta^n < \varepsilon\, |a_n|/2$, we get

$$|f(z)| > \frac{1}{\delta^n} \frac{|a_n|}{2} > \frac{1}{\varepsilon}, \qquad \text{if } |z| > 1/\delta.$$ □

Figure 3.4 shows a phase portrait of the polynomial $f(z) = z^5 - z^4 - z + 1$ on two domains covering the lower hemisphere (left) and the upper hemisphere (right) of $\widehat{\mathbb{C}}$. For the construction of these two images we refer to page 41. The left window is the square $|\mathrm{Re}\, z| < 1.2$, $|\mathrm{Im}\, z| < 1.2$, the four distinctive points where all colors meet are $z_1 = 1, z_2 = \mathrm{i}, z_3 = -1$ and $z_4 = -\mathrm{i}$. These points are also seen in the right window. The point $z_5 = \infty$ is at the center of the square.

Let us compare the phase portrait of f in a neighborhood of the points $z_1, \ldots, z_5$ with the portraits of the power functions z^n near the origin. We observe that, up to a translation and a slight distortion, the phase portrait of f at z_1 looks like the portrait of z^2 at $z = 0$. Similarly, the phase portrait of f at z_2, z_3, z_4

resembles the portrait of z near the origin (for z_2 and z_4 we need to add a rotation), while it mimics z^{-5} at z_5.

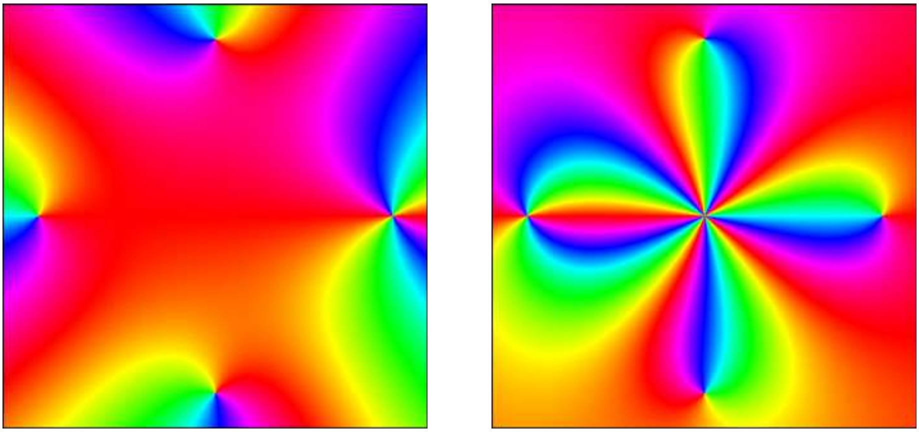


Figure 3.4: Phase portraits of a polynomial of degree 5 on the two hemispheres

In order to analyze these observations we have to do some groundwork. First of all let us study algebraic manipulations with polynomials.

Polynomial Algebra. Let f and g be two polynomials. Then the sum $f+g$, the difference $f-g$, and the product $f \cdot g$ (often denoted by $f\,g$) are polynomials with

$$\deg(f \pm g) \leq \max(\deg f, \deg g), \qquad \deg(f \cdot g) = \deg f + \deg g.$$

After defining $n + (-\infty) := -\infty$, and $\max(n, -\infty) = \max(-\infty, n) := n$ if n is an integer or $n = -\infty$, these relations remain valid if one or both of f and g is the zero polynomial.

The quotient of two polynomials f and g is, in general, not a polynomial. We say that g *divides* f (or f *is divisible by* g), if there exists a polynomial q such that $f = g \cdot q$. In order to study this further, we utilize the technique of *division with remainder.*

Theorem 3.1.2 (Division with Remainder)**.** *Let f and $g \not\equiv 0$ be polynomials. Then there exist uniquely defined polynomials q (the* partial quotient *of f and g) and r (the* remainder*) with $\deg r < \deg g$ such that $f = g\,q + r$.*

Proof. 1. In the first step we prove the existence of q and r. If $\deg f < \deg g$ then we can set $q := 0$ and $r := f$. So let

$$f(z) = a_n\, z^n + \ldots + a_1\, z + a_0, \quad g(z) = b_m\, z^m + \ldots + b_1\, z + b_0$$

with $a_n, b_m \neq 0$ and $n > m$. Setting $q_1(z) := (a_n/b_m)\, z^{n-m}$ and $f_1 := f - g q_1$ we get a polynomial f_1 with $\deg f_1 < \deg f$. If $\deg f_1 < \deg g$ we are done, setting

$r := f_1$, otherwise the procedure is repeated with f replaced by f_1, so that the polynomial $f_2 := f_1 - g q_2$ has degree less than $\deg f_1$. After a finite number of steps we arrive at $f_k := f_{k-1} - g q_k$ with $\deg f_k < \deg g$, and setting $q := (q_1 + \ldots + q_k)$, $r := f_k$ yields $f = g\,q + r$ with $\deg r < \deg g$.

2. In order to prove the uniqueness of the polynomials q and r, we assume that $f = g\,q_1 + r_1 = g\,q_2 + r_2$ with $\deg r_1 < \deg g$ and $\deg r_2 < \deg g$. Then $g\,(q_2 - q_1) = r_2 - r_1$. If $q_2 \neq q_1$, then the left-hand side has degree at least $\deg g$, while the degree of the right-hand side is less than $\deg g$. Consequently $q_2 = q_1$, and then also $r_2 = r_1$. ☐

So, g divides f if and only if the remainder r of the partial division of f by g is the zero polynomial. In what follows we apply this result to study zeros of polynomials.

Lemma 3.1.3. *Let f be a polynomial of degree not less than 1. If $z_0 \in \mathbb{C}$ and $f(z_0) = 0$, then f is divisible by the polynomial $g(z) := z - z_0$.*

Proof. Division with remainder yields $f(z) = (z - z_0)\,q(z) + c$ with a constant c. Inserting $z = z_0$ we get $c = 0$, i.e., $f(z) = (z - z_0)\,q(z)$. ☐

This procedure of factoring off zeros can be repeated. If z_1 is a zero of q, (not necessarily distinct from z_0), then we get $f(z) = (z - z_0)(z - z_1)\,q_1(z)$, and so on. Since the degree of the quotients q_k decreases in every step, this process must come to an end. Collecting all linear factors $(z - z_j)$ with the same value of z_j and renumbering the z_j we finally arrive at

$$f(z) = (z - z_1)^{m_1} \cdot (z - z_2)^{m_2} \cdot \ldots \cdot (z - z_k)^{m_k} \cdot p(z), \tag{3.4}$$

where p is a polynomial of degree $\deg f - (m_1 + m_2 + \ldots + m_k)$ without zeros. It is now easy to see that the zeros of f are precisely the numbers z_j with $j = 1, \ldots, k$. The exponent m_j of the factor $(z - z_j)^{m_j}$ in the representation (3.4) is called the *multiplicity* (or the *order*) of the corresponding zero z_j.

Let us pause here for a moment and discuss how the factorization (3.4) can be utilized to explain some of the observations on the phase portrait we have made earlier. Denoting the product of all but the single factor $(z - z_j)^{m_j}$ on the right-hand side of (3.4) by p_j, we obtain

$$f(z) = (z - z_j)^{m_j}\, p_j(z). \tag{3.5}$$

Since $p_j(z_j) \neq 0$, the phase of p_j is continuous at z_j, such that

$$\psi\big(f(z)\big) = \psi\big((z - z_j)^{m_j}\big) \cdot \psi\big(p_j(z)\big) \sim c \cdot \psi\big((z - z_j)^{m_j}\big), \quad c := \psi\big(p_j(z_j)\big), \tag{3.6}$$

for all z sufficiently close to z_j. So, in a small neighborhood of z_j, the phase portrait of f looks like the phase portrait of z^{m_j} at $z = 0$, shifted to z_j, and rotated by an angle of

$$\varphi_j = -(1/m_j)\, \arg p_j(z_j). \tag{3.7}$$

The rotation angle in (3.7) is probably not what one would expect at the first glance. To verify that it is correct, one should take into account that the phase portrait is generated by a pull-back operation, which explains the negative sign and the factor $1/m_j$.

In order to interpret the phase portrait of f at $z = \infty$, we use the representation

$$f(z) = z^n \left(a_n + \frac{a_{n-1}}{z} + \ldots + \frac{a_1}{z^{n-1}} + \frac{a_0}{z^n} \right), \quad z \neq 0, \tag{3.8}$$

and obtain

$$\psi\big(f(z)\big) \sim c \cdot \psi\big(z^n\big), \quad c := \psi\big(a_n\big). \tag{3.9}$$

Extremal Values. Now one crucial question remains: how many zeros does a polynomial of degree n have? The representation (3.4) tells us that this number is $\deg f - \deg p$, counting all zeros according to their multiplicities. So, what can be said about the polynomial p? One approach to answer these questions is the following result on local extremal values of polynomials.

Theorem 3.1.4 (Maximum / Minimum Principle for Polynomials)**.**

(i) *If the modulus $|f|$ of a polynomial f has a local maximum at some point $z_0 \in \mathbb{C}$, then f is constant.*

(ii) *If f is a polynomial of degree at least one, and $|f|$ has a local minimum at some point $z_0 \in \mathbb{C}$, then $f(z_0) = 0$.*

Proof. 1. We assume that $n := \deg f \geq 1$ and pick any $z_0 \in \mathbb{C}$ with $f(z_0) \neq 0$. Then both assertions follow, once we have shown that $|f|$ attains neither a local maximum nor a local minimum at z_0.

2. Inserting $z = z_0 + (z - z_0)$ into (3.1) we can represent f as a polynomial in $z - z_0$,

$$f(z) = c_0 + c_1\,(z - z_0) + c_2\,(z - z_0)^2 + \ldots + c_n\,(z - z_0)^n, \tag{3.10}$$

where $c_0 = f(z_0) \neq 0$. Since f is not a constant, not all coefficients $c_1, \ldots, c_n$ can vanish. Consequently we can rewrite (3.10) as

$$f(z) = c_0 + c_k\,(z - z_0)^k\,\big(1 + g(z)\big),$$

where g is a polynomial with $g(z_0) = 0$. Since g is continuous, there exists a positive δ such that $|g(z)| < 1/2$ for all z with $|z - z_0| < \delta$.

3. Now let ε be an arbitrary positive number. In order to prove (i), we choose ω as a kth root of c_0/c_k, that is, $\omega^k = c_0/c_k$. Setting $z = z_0 + r\,\omega$ with $r \in \mathbb{R}$ and $0 < r < \min(1, \varepsilon/|\omega|, \delta/|\omega|)$, we have $|z - z_0| < \min(\varepsilon, \delta)$, so that $|g(z)| < 1/2$. Then

$$f(z) = c_0\,\big(1 + r^k + r^k\,g(z)\big),$$

and, by the triangle inequality,

$$|f(z)| \geq |c_0|\,(1 + r^k - (1/2)r^k) > |c_0| = f(z_0).$$

In order to prove (ii) we choose ω such that $\omega^k = -c_0/c_k$, and get analogously

$$|f(z)| = \left|c_0\left(1 - r^k - r^k\, g(z)\right)\right| \le |c_0|\left(1 - r^k + (1/2)r^k\right) < |c_0| = f(z_0). \qquad \square$$

Corollary 3.1.5. *Any polynomial f of degree at least one has a zero z_0 in $\mathbb{C}$.*

Proof. We consider the closed disk $K := \{z \in \mathbb{C} : |z| \le R\}$. Since $f(\infty) = \infty$ and $f : \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ is continuous, the radius R can be chosen so large that $|f(z)| > |f(0)|$ for all z with $|z| \ge R$. The function $|f|$ is continuous on the compact set K. By Weierstrass' theorem, it must attain a minimum at some point $z_0 \in K$. Since $|f(z_0)| \le |f(0)|$, we infer from the first step that z_0 must be an inner point of K. By Theorem 3.1.4 (ii), this can only happen if $f(z_0) = 0$. $\square$

Zeros. The following celebrated theorem on zeros of polynomials was first proved by different techniques (and according to modern standards still with some gaps) by Carl Friedrich Gauss in 1799.

Theorem 3.1.6 (Fundamental Theorem of Algebra)**.** *Any polynomial f of degree $n \ge 1$ has exactly n complex zeros. More precisely, denoting the different zeros of f by $z_1, z_2, \ldots, z_k$, their multiplicities $m_1, m_2, \ldots, m_k$ satisfy $m_1 + m_2 + \ldots + m_k = n$, and f admits the* canonical factorization

$$f(z) = c\,(z - z_1)^{m_1} \cdot (z - z_2)^{m_2} \cdot \ldots \cdot (z - z_k)^{m_k}, \quad c \in \mathbb{C} \setminus \{0\}.$$

Proof. Factoring off all zeros of f (allowing also for the case when f has no zeros at all) we represent f in the form (3.4), with a polynomial p having no zeros, and $c \in \mathbb{C} \setminus \{0\}$. By Corollary 3.1.5, p must be constant. $\square$

The next result is an immediate consequence of Theorem 3.1.6.

Corollary 3.1.7 (Identity Theorem for Polynomials)**.** *Let $n \ge 0$ and assume that $f(z) = a_0 + a_1\, z + \ldots + a_n\, z^n$ and $g(z) = b_0 + b_1\, z + \ldots + b_n\, z^n$ are polynomials of degree not exceeding n. If $f(z_j) = g(z_j)$ for $n+1$ distinct points $z_0, z_1, \ldots, z_n \in \mathbb{C}$, then $a_j = b_j$ for $j = 0, \ldots, n$. In particular $f(z) = g(z)$ for all $z \in \mathbb{C}$.*

Proof. If $a_j \ne b_j$ for some $j \in \{1, \ldots, n\}$, then $\deg(f - g) \ge 1$, which contradicts Theorem 3.1.6. Thus $f(z) - g(z) = a_0 - b_0$, and then $a_0 = b_0$. $\square$

An Example. The polynomial $f(z) = z^5 - z^4 - z + 1$ has zeros at $z_1 = 1$, $z_2 = \mathrm{i}$, $z_3 = -1$, and $z_4 = -\mathrm{i}$. While z_1 has multiplicity two, z_2, z_3, and z_4 have multiplicity one, so that f has the canonical factorization

$$f(z) = (z - 1)^2\,(z + 1)(z^2 + 1).$$

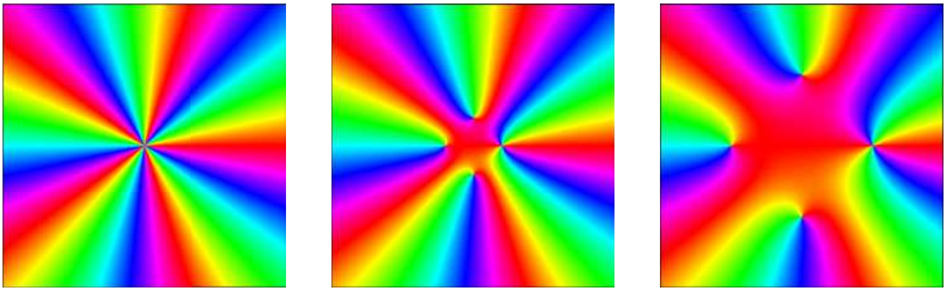


Figure 3.5: Zooming in to the phase portrait of the polynomial $f(z) = z^5 - z^4 - z + 1$

Let us look again at the phase portrait of this function. The three windows of Figure 3.5 are centered at the origin and have side lengths 100, 10 and 4, respectively. In the window on the left with the smallest zoom factor all details are indiscernible, and we just see a phase portrait similar to that of the function z^5, which is a consequence of (3.9). Moving closer, the zeros show up.

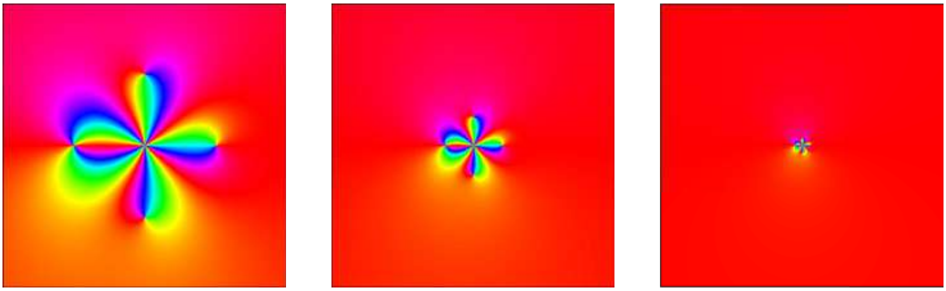


Figure 3.6: Zooming out at infinity of the phase portrait of the polynomial f

An Alternative Proof? Looking at these pictures for a while with attention focused to one color, say yellow, we observe that any zero is linked to the point at infinity by an isochromatic line of that color. Or conversely, each of the five yellow lines emerging at $z = \infty$ ends up at a zero. This can be seen best in Figure 3.6 (left) which depicts a neighborhood of the point at infinity. Note that two yellow lines end up at the double zero $z_1 = 1$.

If this observation could be confirmed for general polynomials, it would give an intuitive interpretation of the fundamental theorem of algebra: the number of yellow lines emerging at infinity is equal to the degree of the polynomial, every line ends up at a zero, and the number of lines terminating at a zero is equal to its multiplicity. We shall convert these observations into actual results when we study *phase diagrams* in Volume 2.

Equipped with these tools we can go even further: fix a zero z_0 of f and consider all isochromatic lines (of all colors) ending up at z_0. These lines fill some region B in the complex plane. If we pick a point in B and let it travel along the corresponding isochromatic line it will finally end up at z_0. So, B is the *basin of attraction* of the zero z_0. Considering *all* zeros of f, their basins of attraction (almost) cover the complex plane.

Interpolation. Corollary 3.1.7 tells us that any polynomial f of degree n is uniquely determined by its values $w_j = f(z_j)$ at $n+1$ distinct points z_j. Conversely, we may ask if the numbers w_j can be prescribed arbitrarily. Indeed there is a positive answer and the *interpolation problem* of finding a polynomial f from its values w_j at z_j can even be solved algorithmically. Here we consider only a special situation, where the *coefficients* of the interpolating polynomial f are given by a beautiful formula.[1]

Lemma 3.1.8. *Let $f = a_0 + a_1\,z + a_2\,z^2 + \ldots + a_n\,z^n$ be a polynomial of degree at most n, and let ω denote the principal $(n+1)$ th root of unity,*

$$\omega := \cos\frac{2\pi}{n+1} + \mathrm{i}\,\sin\frac{2\pi}{n+1}.$$

Then

$$a_k = \frac{1}{n+1}\sum_{j=0}^{n}\omega^{-jk}\,f\big(\omega^j\big), \quad k = 0, 1, \ldots, n. \tag{3.11}$$

Proof. First of all we observe that for *any* $(n+1)$ th root of unity ω with $\omega \neq 1$

$$1 + \omega + \omega^2 + \ldots + \omega^n = \frac{\omega^{n+1} - 1}{\omega - 1} = 0.$$

Consequently, summing up the equations

$$\omega^{-jk}\,f\big(\omega^j\big) = \sum_{i=0}^{n} a_i\,\omega^{j(i-k)} = \sum_{i=0}^{n} a_i\,\big(\omega^{i-k}\big)^j$$

for $j = 0, \ldots, n$ all terms get cancelled except those with $i = k$, in which case we have $\big(\omega^{i-k}\big)^j = 1$, and hence the sum equals $(n+1)\,a_k$. □

If the polynomial is evaluated at $n+1$ uniformly distributed points on a circle with radius r and center 0, one can use a scaled version of formula (3.11),

$$r^k\,a_k = \frac{1}{n+1}\sum_{j=0}^{n}\omega^{-jk}\,f\big(r\,\omega^j\big), \quad k = 0, 1, \ldots, n. \tag{3.12}$$

---

[1] The experienced reader will notice the connection with the discrete Fourier transform.

Rational Functions. In the next step of constructing complex functions we also allow *division*, and enter the realm of *rational functions*. These functions are quotients of polynomials

$$f(z) = \frac{p(z)}{q(z)} = \frac{a_0 + a_1\, z + a_2\, z^2 + \ldots + a_n\, z^n}{b_0 + b_1\, z + b_2\, z^2 + \ldots + b_m\, z^m}, \tag{3.13}$$

with $q \not\equiv 0$. The natural domain set of f consists of all points $z \in \mathbb{C}$ where q does not vanish. If the *numerator* p and the *denominator* q have a common zero z_0 with multiplicities m_p and m_q, respectively, then the common factor $(z - z_0)^m$ with $m := \min(m_p, m_q)$ in the canonical factorization of p and q can be cancelled in the quotient p/q. If this is done for all common zeros of p and q we get the *irreducible* representation of a rational function. The *degree* of a rational function is defined as $\deg f = \max(\deg p, \deg q)$, where $f = p/q$ is the irreducible representation of f.

In order to extend a rational function to all of $\widehat{\mathbb{C}}$ we represent it in the irreducible form $f = p/q$. In doing so we may have already extended its domain of definition. Further we define $f(z) := \infty$ for all $z \in \mathbb{C}$ with $q(z) = 0$. Finally we set

$$f(\infty) := \begin{cases} \infty & \text{if} \quad \deg p > \deg q \\ a_n/b_n & \text{if} \quad \deg p = \deg q = n \\ 0 & \text{if} \quad \deg p < \deg q. \end{cases} \tag{3.14}$$

These definitions are natural since they guarantee the continuity of the extended function.

Theorem 3.1.9. *Any rational function f is a continuous mapping $f : \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$.*

Proof. We may assume that $f = p/q$ is the irreducible form of f. Then f is continuous at all points $z_0 \in \mathbb{C}$ with $q(z_0) \neq 0$. If $z_0 \in \mathbb{C}$ and $q(z_0) = 0$ then $f(z_0) := \infty = \lim_{z\to\infty} f(z)$. Finally, the definition of $f(\infty)$ is such that it coincides with $\lim_{z\to\infty} f(z)$. □

Zeros and Poles. The zeros of the extended function f on $\mathbb{C}$ coincide with the zeros of p. If $z_0 \in \mathbb{C}$ is a zero of p with multiplicity k then $f(z) \sim c\,(z - z_0)^k$ at z_0 with $c \neq 0$,[2] and z_0 is said to be a *zero* of f *with multiplicity* k. Analogously, if $z_0 \in \mathbb{C}$ is a zero of q with multiplicity k, then $f(z) \sim c\,(z - z_0)^{-k}$, and z_0 is called a *pole* of f *with multiplicity* k. Define $k := \deg p - \deg q$. If $k < 0$, the point at infinity is said to be a *zero with multiplicity* k, if $k > 0$ then f has a *pole of multiplicity* $|k|$ *at infinity.*

The phase portrait had taken note of all this all the while and has done the cancellation automatically. Hence it looks similar to the portraits of the corresponding power functions $c\,(z - z_0)^k$ in a neighborhood of poles and zeros.

---

[2] For the definition of the symbol $\sim$ see page 45.

Zero-Pole-Cancellation. It should be mentioned that, in contrast to zeros of polynomials, which always have a "long range effect", zeros and poles at small distances can *almost cancel* each other, so that they may be invisible in the phase portrait. This effect is demonstrated in Figure 3.7 for a zero-pole pair approaching the center of the square depicted.

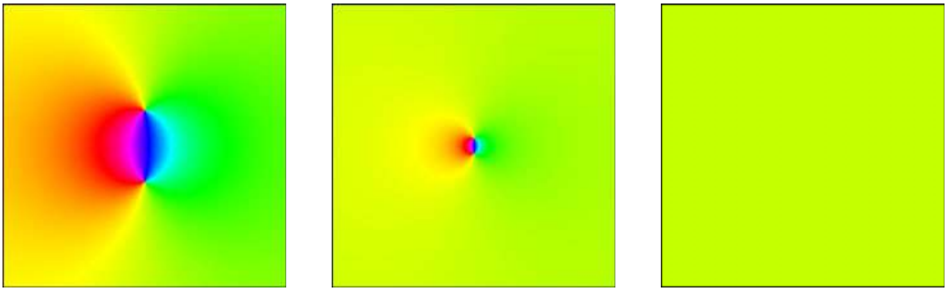


Figure 3.7: An almost zero-pole-cancellation may be invisible

A-Values. That the above extension of a rational function f to all of $\widehat{\mathbb{C}}$ was reasonable becomes clear when we consider the *a-values* of f. For $a \in \mathbb{C}$ these are the zeros of $f - a$, the ∞-values are just the poles of f. In the following theorem all a-values are counted according to the multiplicity of zeros of $f - a$.

Theorem 3.1.10. *For any $a \in \widehat{\mathbb{C}}$ the number of a-values of a non-constant rational function f is equal to* $\deg f$.

Proof. Let $f = p/q$ be an irreducible representation with $\deg p = m$, $\deg q = n$, and let $k := \deg f = \max(m, n)$. We distinguish different cases.

1. If $a \in \mathbb{C}$, then the polynomial $g := aq - p$ has degree $j \le k$, and g has exactly j zeros $z_1, \dots z_j$ in $\mathbb{C}$. Since p and q are have no common zeros, none of the z_i can be a zero of q, so $z_1, \dots z_j$ are precisely the finite zeros of $p/q - a$. Further ∞ is a zero of $p/q - a = (p - aq)/q$ with multiplicity $k - j$.

2. If $a = \infty$, then p/q has exactly $n = \deg q$ finite poles. If $n \ge m$, then ∞ is not a pole of f, otherwise it is a pole of order $m - n$. In both cases the total number of poles equals $\deg f$. □

In geometric language, a rational function $z \mapsto w = f(z)$ of degree n induces a *n-fold covering* of the w sphere by the image of the z sphere.

Partial Fraction Decomposition. Besides the *multiplicative normal form* of rational functions, which we immediately obtain from the canonical factorizations of the polynomials p and q, there is also an *additive normal form.*

Theorem 3.1.11 (Partial Fraction Decomposition)**.** *Let $f = p/q \not\equiv 0$ be an irreducible representation of a rational function f, and denote by $z_1, \dots, z_k$ the zeros of q with the corresponding multiplicities $m_1, \dots, m_k$. Then f admits a partial* *fraction decomposition* $f = f_0 + f_1 + \ldots + f_k$, *where* f_0 *is a polynomial of degree* $\deg p - \deg q$ *if* $\deg p \geq \deg q$ *and* $f_0 = 0$ *otherwise, and*

$$f_j(z) = \frac{a_{j1}}{(z-z_j)^{m_j}} + \frac{a_{j2}}{(z-z_j)^{m_j-1}} + \ldots + \frac{a_{jm_j}}{z-z_j}, \quad j = 1, \ldots, k.$$

Before we give a proof, let us look at the phase portrait to illustrate the main idea. If f has a pole of order m at z_0, then its phase portrait resembles locally the portrait of $1/(z-z_0)^m$, rotated by a certain angle. This rotation corresponds to multiplication by some factor a. So we try to find a constant a such that $a/(z-z_0)^m$ is the best fit function to f in a neighborhood of z_0. If a is chosen correctly, the poles of f and of $a/(z-z_0)^m$ at z_0 partially compensate each other, so that $f_1(z) = f(z) - a/(z-z_0)^m$ has a pole of order not exceeding $m-1$. Then the same process is repeated for the new function f_1, and so on, until there are no more poles at all. Figure 3.8 illustrates two steps of this procedure for a pole of second order located at the center of the square. Note that the two other poles in the domain remain unaffected.

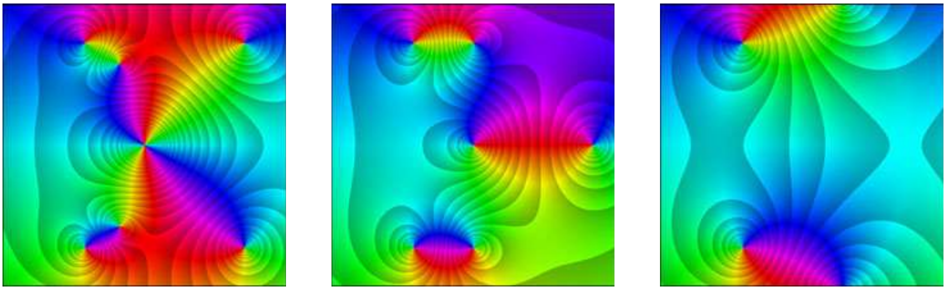


Figure 3.8: Partial fraction decomposition at a pole of second order

Proof. If $\deg q = 0$, then f is a polynomial. If $\deg p \geq \deg q \geq 1$, partial division of p by q yields $p = q\, f_0 + p_1$, such that $f = f_0 + p_1/q$ is the sum of a polynomial f_0 and a rational function p_1/q with $\deg p_1 < \deg q$. Therefore it suffices to consider rational functions p_1/q with $\deg p_1 < \deg q$.

If $\deg q = 1$, the result is obvious. Now we argue by mathematical induction. Assuming that, for some positive integer n, the assertion holds for all rational functions p_1/q with $\deg p_1 < \deg q \leq n$, we wish to show that it also holds when $\deg p_1 < \deg q = m = n + 1$. Since z_1 is a zero of q with multiplicity m_1, the polynomial q can be represented as

$$q(z) = (z - z_1)\, q_1(z) = (z - z_1)^{m_1}\, g_1(z)$$

with polynomials q_1 and g_1 satisfying $g_1(z_1) \neq 0$. Setting $a_{11} := p_1(z_1)/g_1(z_1)$, we obtain

$$\frac{p_1(z)}{q(z)} - \frac{a_{11}}{(z-z_1)^{m_1}} = \frac{p_1(z) - a_{11}\, g_1(z)}{q(z)} = \frac{p_1(z) - a_{11}\, g_1(z)}{(z-z_1)\, q_1(z)}. \tag{3.15}$$

By definition of a_{11}, the numerator $p_2(z) := p_1(z) - a_{11}\, g_1(z)$ vanishes at z_1, and thus the factor $z - z_1$ on the right-hand side of (3.15) gets cancelled. Hence

$$\frac{p_1(z)}{q(z)} - \frac{a_{11}}{(z - z_1)^{m_1}} = \frac{p_2(z)}{q_1(z)},$$

with $\deg p_2 < \deg q_1 \leq m - 1 = n$. The zeros of q_1 and their multiplicities coincide with those of q, except for z_1 which now has multiplicity $m_1 - 1$. By assumption, the rational function p_2/q_1 admits the desired decomposition, which completes the proof of the theorem. $\square$

3.2 Power Series

Using merely the four basic arithmetic operations one cannot go beyond the class of rational functions. To obtain new functions we need to use other constructive methods, for instance *limit processes.*

An Example. To get the flavor of such an approach, we consider an infinite sequence of complex numbers $a_0, a_1, a_2, \ldots$ and define the sequence of polynomials

$$f_0(z) = a_0, \quad f_1(z) = a_0 + a_1\, z, \quad f_2(z) = a_0 + a_1\, z + a_2\, z^2, \quad \ldots\ .$$

What happens with these functions f_n when their degrees increase? To have a concrete example, take all coefficients equal to one, $a_0 = a_1 = a_2 = \ldots = 1$. Figure 3.9 shows the phase portraits of the corresponding polynomials of degree 20, 60 and 150.

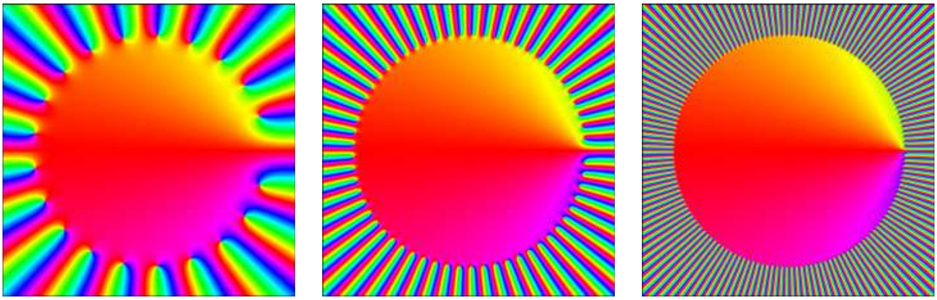


Figure 3.9: Phase portraits of the polynomials f_{20}, f_{60} and f_{150}

We observe that the behavior of these phase portraits is quite different, depending on the region which we observe. There is a disk-like region where the coloring does not change much when n is increased and becomes stable for sufficiently large n. On the other hand, in the outer region the stripes change very rapidly, and no definite color can be assigned to a point in that region for large n. Of course, these three phase portraits say nothing about the *actual* behavior of the functions f_n.

To see that this observation is not accidental we need some more mathematical input.

However, the example at hand can be easily explained. The polynomial f_n is the sum of a finite *geometric series*, which can be computed explicitly. If $z \neq 1$ we have

$$f_n(z) = 1 + z + z^2 + \ldots + z^n = \frac{z^{n+1} - 1}{z - 1},$$

and $f_n(1) = n$. Thus the sequence $\big(f_n(z)\big)$ converges to the limit $1/(1-z)$ if $|z| < 1$, while it diverges for all $z \in \mathbb{C}$ with $|z| \geq 1$.

Power Series. Now, let us explore the general case of the previous example. First of all we modify the setting a little in order to get a bit more flexibility. Fix $z_0 \in \mathbb{C}$, and let $(a_n)_{n=0}^{\infty}$ be a sequence of complex numbers. Then the function series

$$\sum_{k=0}^{\infty} a_k\,(z - z_0)^k = a_0 + a_1\,(z - z_0) + a_2\,(z - z_0)^2 + \ldots \tag{3.16}$$

is called a *power series*. Here and in the following we set $(z - z_0)^0 := 1$ for all z, otherwise the first summand becomes indeterminate if $z = z_0$. The numbers a_k are the *coefficients* of the series and the point z_0 is referred to as its *center*. The *partial sums* of the series are the polynomials f_n defined by

$$f_n(z) := a_0 + a_1\,(z - z_0) + a_2\,(z - z_0)^2 + \ldots + a_n\,(z - z_0)^n. \tag{3.17}$$

We say that the power series *converges at the point* z, if the sequence of complex numbers $f_n(z)$ converges, and its limit is called the *sum* of the power series (3.16) at z. Otherwise the power series (3.16) is said to be *divergent* at z. If the series converges at all points z in a set D, its sum $f(z)$ defines a complex function $f : D \to \mathbb{C},\ z \mapsto f(z)$, which we write as

$$f(z) := \sum_{k=0}^{\infty} a_k\,(z - z_0)^k = a_0 + a_1\,(z - z_0) + a_2\,(z - z_0)^2 + \ldots\,. \tag{3.18}$$

The Disk of Convergence. The following result shows that the behavior of the power series in the above example is typical: the series converges for all z within a *disk of convergence* centered at z_0, outside the closure of that disk the series diverges. We denote by α the limit superior of $\sqrt[k]{|a_k|}$ and set $R := 1/\alpha$ with the usual conventions $1/0 := \infty$ and $1/\infty := 0$,

$$R := 1 \Big/ \limsup_{k \to \infty} \sqrt[k]{|a_k|}. \tag{3.19}$$

Theorem 3.2.1 (Cauchy-Hadamard)**.** *The power series* (3.16) *converges for all* $z \in \mathbb{C}$ *with* $|z - z_0| < R$ *and diverges for all* $z \in \mathbb{C}$ *with* $|z - z_0| > R$*. For any* r *with* $0 < r < R$ *the series converges absolutely and uniformly on the closed disk* $K_r := \{z \in \mathbb{C} : |z - z_0| \leq r\}$.

Proof. In the following we assume that $0 < R < \infty$, the cases $R = 0$ and $R = \infty$ require only minor modifications and are left as exercises.

1. Let $\alpha_k = \sqrt[k]{|a_k|}$. By assumption $0 < \alpha := \limsup \alpha_k < \infty$, and thus the limit superior α is the uniquely determined real number with the following property: for any positive ε there exists an infinite set of indices k with $\alpha_k > \alpha - \varepsilon$, while $\alpha_k > \alpha + \varepsilon$ holds for at most finitely many k.

2. Fix a positive number r with $r < R$, set $q := \sqrt{r/R}$, and suppose that $z \in \mathbb{C}$ satisfies $|z - z_0| \le r$. Then, by the characterization of the limit superior in the first step, there exists a natural number k_0 such that for all $k \ge k_0$

$$\sqrt[k]{|a_k|}\,|z - z_0| \le \frac{\alpha}{q}\,|z - z_0| = \frac{|z - z_0|}{qR} \le \frac{r}{qR} = q < 1. \tag{3.20}$$

Thus the convergent geometric series $\sum q^k$ majorizes the power series (3.16), which implies that the latter converges absolutely and uniformly for all z with $|z-z_0| \le r$. Since r can be chosen arbitrarily close to R the first assertion follows.

3. If $z \in \mathbb{C}$ satisfies $|z - z_0| > R$, then $q := R/|z - z_0| < 1$. By Step 1 there exist infinitely many indices k such that $\sqrt[k]{|a_k|} \ge q/R$, and hence $\left|a_k\,(z-z_0)^k\right| \ge 1$. Consequently the power series cannot converge at z because its summands do not tend to zero. □

In what follows we say that a power series *converges* if its *radius of convergence* R is not zero. Then the *disk of convergence*

$$D := \{z \in \mathbb{C} : |z - z_0| < R\}$$

is either an open disk in $\mathbb{C}$, or the full complex plane $\mathbb{C}$. Note that we do not consider the convergence of power series at $z = \infty$.

Convergence on the Boundary. We mention that no general statement can be made about convergence for points on the boundary of the disk of convergence. The three power series

$$\sum_{k=0}^{\infty} z^k, \qquad \sum_{k=1}^{\infty} \frac{z^k}{k}, \qquad \sum_{k=1}^{\infty} \frac{z^k}{k^2} \tag{3.21}$$

all have radius of convergence $R = 1$. The first one converges for no z on the unit circle $\mathbb{T}$, the second one converges for $z = -1$ (Leibniz series) but diverges for $z = 1$ (harmonic series). The third one converges for all $z \in \mathbb{T}$ because the series $\sum 1/k^2$ is a majorant.

Figure 3.10 shows phase portraits of the partial sums f_{200} of these series. Though we see some differences in the behavior of these polynomials (of degree 200) near the exterior side of the boundary of the disk of convergence (shown in black), it should be clear that this says nothing about the actual convergence of the series.

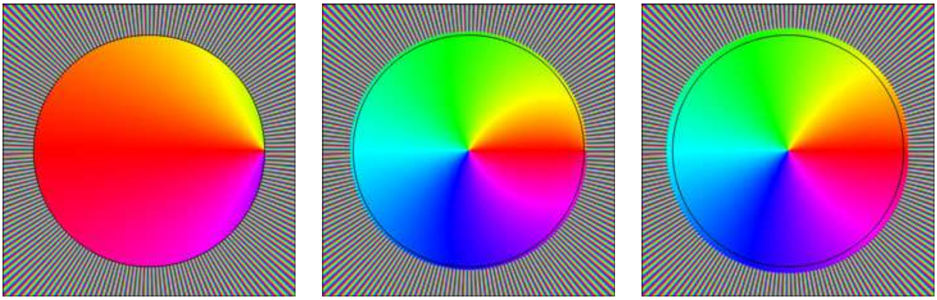


Figure 3.10: Phase portraits of the partial sums f_{200} of the power series (3.21)

Alternative Characterizations of R. Analyzing the proof of the Cauchy-Hadamard theorem we obtain an alternative characterization of the radius of convergence which avoids the limit superior: R is the supremum of all $r \geq 0$ for which the sequence $\big(r^k|a_k|\big)$ is *bounded*. The following reformulation of this statement is even more convenient in applications.

Lemma 3.2.2 (Abel-Weierstrass). *Let R be the radius of convergence of the power series (3.16).*

(i) *If $0 \leq r < R$, then there exists a constant c such that for all $k \in \mathbb{N}$*

$$r^k|a_k| \leq c. \tag{3.22}$$

(ii) *If there exist positive numbers r and c such that (3.22) holds for all sufficiently large $k \in \mathbb{N}$, then $R \geq r$.*

The inequality (3.22) is also known as *Cauchy's estimate*.

Proof. 1. Setting $z = z_0 + r$ in (3.20) we get the desired estimate with $c = 1$ for all sufficiently large k. By increasing c, if necessary, we can also capture the finite number of remaining a_k. The value of the constant c will be specified later (see (3.27)).

2. To prove the second result we remark that the estimate (3.22) implies $\sqrt[k]{|a_k|} \leq \sqrt[k]{|c|}/r \to 1/r$, and apply the Cauchy-Hadamard formula. □

Of course, the limit superior in the Cauchy-Hadamard formula (3.19) can also be replaced by the ordinary limit if the latter exists. In many applications it is easier to use the *ratio test*, which yields that

$$R = \lim_{k\to\infty} \frac{|a_k|}{|a_{k+1}|},$$

provided that this (finite or infinite) limit exists. Let us look at an example next.

Example 3.2.1 (Binomial Series). For $\alpha \in \mathbb{R}$ and $k \in \mathbb{N}$ the *generalized binomial coefficients* are given by

$$\binom{\alpha}{0} := 1, \qquad \binom{\alpha}{k} := \frac{\alpha \cdot (\alpha-1) \cdot \ldots \cdot (\alpha-k+1)}{k \cdot (k-1) \cdot \ldots \cdot 2 \cdot 1}.$$

The *binomial series* with center $z_0 = 1$ and parameter $\alpha \in \mathbb{R}$ is the power series

$$\sum_{k=0}^{\infty} \binom{\alpha}{k} (z-1)^k. \tag{3.23}$$

If α is a natural number, then the series has only a finite number of non-zero summands. Otherwise its radius of convergence can be determined by the ratio test,

$$R = \lim_{k\to\infty} \frac{|a_k|}{|a_{k+1}|} = \lim_{k\to\infty} \left|\binom{\alpha}{k}\right| \Big/ \left|\binom{\alpha}{k+1}\right| = \lim_{k\to\infty} \left|\frac{k+1}{\alpha-k}\right| = 1.$$

If z is real and $0 < z < 2$, the sum of this series is the *fractional power function* $f(z) = z^\alpha$. It can be shown that for rational $\alpha = m/n$ the sum f of the binomial series satisfies $\big[f(z)\big]^n = z^m$ for all z in the unit disk.

Figure 3.11 depicts phase portraits of the sums of the binomial series for $\alpha = -2, 0.5$, and 7.5, respectively. The region outside the disk of convergence is shown in gray.

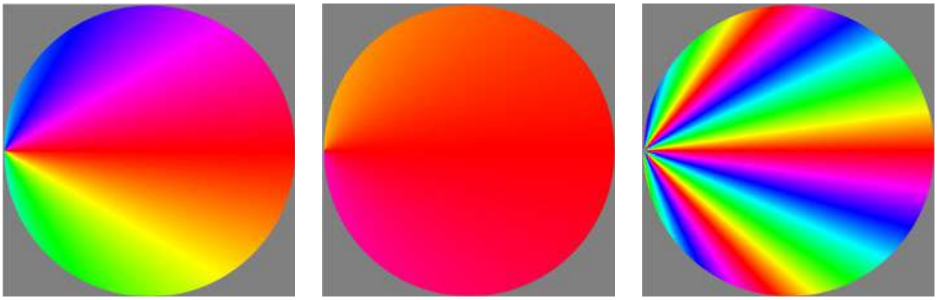


Figure 3.11: Phase portraits of the sums of three binomial series

Power Series of Real Functions. Power series are indispensable tools for extending functions from the real line to a complex domain. Like the fractional power function in the preceding example, many functions of a real variable x can (locally) be represented by a convergent series

$$f(x) = a_0 + a_1 (x-x_0) + a_2 (x-x_0)^2 + \ldots + a_k (x-x_0)^k + \ldots . \tag{3.24}$$

Substituting the real variable x by a complex variable z, we obtain a complex power series centered at the point $x_0 \in \mathbb{R}$. If the real series converges for some

$x_1 \in \mathbb{R}$, then the complex series converges for all z with $|z - x_0| < r := |x_1 - x_0|$. This allows us to *extend* the function f to the disk $D_r(x_0)$ in the complex plane.

Example 3.2.2 (The Logarithm). The power series with center $z_0 = 0$ and coefficients $a_k := (-1)^{k+1}/k$,

$$\sum_{k=1}^{\infty} (-1)^{k+1} \frac{z^k}{k} = z - \frac{z^2}{2} + \frac{z^3}{3} - \frac{z^4}{4} \pm \ldots, \tag{3.25}$$

has radius of convergence one,

$$R = \lim_{k\to\infty} \frac{|a_k|}{|a_{k+1}|} = \lim_{k\to\infty} \frac{k+1}{k} = 1.$$

If z is real and $-1 < z < 1$, the sum of the series (3.25) coincides with the real logarithm function $\log(z+1)$. Thus (3.25) defines a function which extends $\log(z+1)$ from the interval $(-1, 1)$ to the complex unit disk.

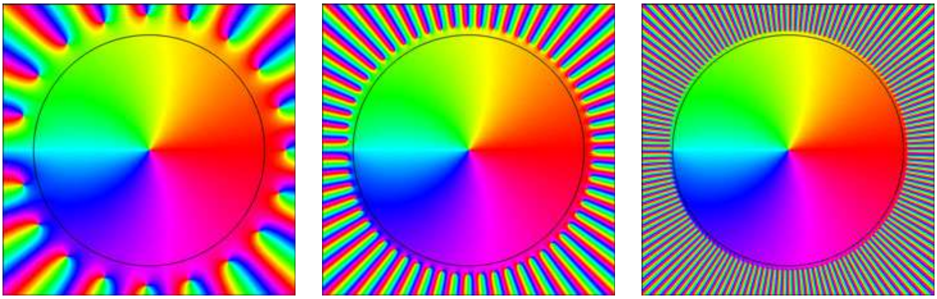


Figure 3.12: Phase portraits of the polynomials f_{20}, f_{60}, and f_{150}

Figure 3.12 shows the phase portraits of the partial sums f_{20}, f_{60} and f_{150}. The black circles indicate the position of the unit circle which is the boundary of the disk of convergence.

Example 3.2.3 (The Exponential Function). The power series of the (real) exponential function at $x_0 = 0$ has the coefficients $a_k = 1/k\,!$, with $k\,! := 1 \cdot 2 \cdot \ldots \cdot k$, and converges on the real axis. Thus an extension e^z of this function to the entire complex plane is given by the power series

$$\mathrm{e}^z = \sum_{k=0}^{\infty} \frac{z^k}{k\,!} = 1 + \frac{z}{1\,!} + \frac{z^2}{2\,!} + \ldots + \frac{z^k}{k\,!} + \ldots, \qquad z \in \mathbb{C}. \tag{3.26}$$

Instead of e^z we shall also sometimes write $\exp z$. Figure 3.13 shows phase portraits of the partial sums f_{10}, f_{20}, and f_{50} of the exponential series (3.26) in the square given by $|\operatorname{Re} z| \le 10$, $|\operatorname{Im} z| \le 10$. The last picture practically coincides with the phase portrait of e^z.

The phase of e^z seems to be constant on lines parallel to the real axis, i.e., $\psi(e^z)$ depends only on $\operatorname{Im} z$. This cannot be easily seen from the definition of the exponential function in (3.26), but it will soon be verified.

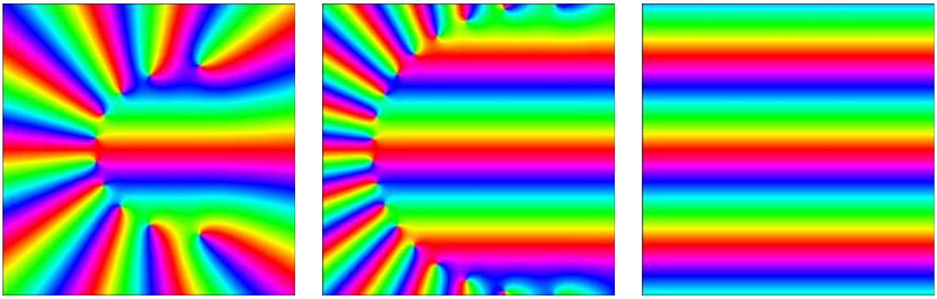


Figure 3.13: The partial sums f_{10}, f_{20}, and f_{50} of the exponential series

Zeros of Partial Sums. Looking at the phase portraits depicted in Figures 3.9, 3.10 and 3.12 one gets the impression that zeros of the partial sums have a tendency to cluster along the boundary of the disk of convergence. Indeed this observation has a sound background.

A theorem which dates back to Robert Jentzsch[3] in 1914 (published in [28], see also Section 7.8 of Titchmarsh [65]) states that the zeros of a power series (3.16) with a positive finite convergence radius R *cluster at every point* z with $|z-z_0| = R$. The reader interested in a contemporary proof is referred to Luh [37].

Since the power series of the exponential function converges in the entire plane the zeros of its partial sums must behave differently. An interesting structure appears when we consider the phase portraits of $f_n(z/n)$. Note that changing the argument from z to z/n effects a scaling of the complex plane. Figure 3.14 shows three such phase portraits with $n = 20, 40$ and 60.

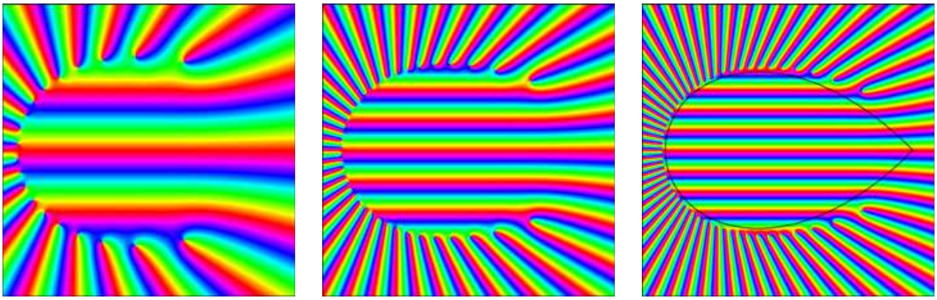


Figure 3.14: Phase portraits of $f_n(z/n)$ of the exponential series with $n = 20, 40, 60$

---

[3] An account of Jentzsch's tragic life is given by Duren, Herbig and Khavinson [13].

A famous result obtained by Gábor Szegő [64] in 1924 confirms the observation that the zeros of the scaled polynomials cluster along a curve. This *Szegő curve* is formed by all points with $\operatorname{Re} z \le 1$ satisfying the equation $\left|z\,\mathrm{e}^{1-z}\right| = 1$ and is depicted in Figure 3.14 (right). All zeros lie in the exterior of the Szegő curve and any point of that curve is a cluster point of zeros.

Properties of the Sums. The sum of a convergent power series defines a function on the disk of convergence D. The properties of such functions will be studied next.

Theorem 3.2.3 (Continuity of the Sum). *The sum of a power series is a continuous function in the disk of convergence.*

Proof. The result follows immediately from Theorem 3.2.1 and Theorem 2.6.2. ☐

Using Theorem 3.2.3 we can now specify the constant in the coefficient estimate of Lemma 3.2.2.

Lemma 3.2.4 (Cauchy Estimate). *Let R be the radius of convergence of the power series (3.18). Then for all r with $0 < r < R$,*

$$|a_k| \le r^{-k} \max_{|z-z_0|=r} |f(z)|, \qquad k = 0, 1, 2, \ldots. \tag{3.27}$$

Proof. The partial sum f_n of the power series (3.18) is a polynomial of degree at most n, and hence the coefficient formula (3.12) tell us that

$$r^k\,|a_k| \le \frac{1}{n+1} \sum_{j=0}^{n} \left|f_n\left(r\omega^j\right)\right| \le \max_{|z|=r} |f_n(z)|.$$

Now the assertion follows, since f_n converges uniformly to f on the disk $|z| \le r$. ☐

Analytic Functions. So far power series have been the primary objects of our consideration, and functions emerged as their sums. Now we reverse the situation: given a complex function f on some domain set $D \subset \widehat{\mathbb{C}}$, we ask whether it can be locally represented as a sum of a power series.

Definition 3.2.5. A function $f : D \to \widehat{\mathbb{C}}$ is said to be *analytic at the point* z_0 in $\mathbb{C}$, if there exists a non-void disk $D_r(z_0) \subset D$ centered at z_0 such that the restriction of f to $D_r(z_0)$ is the sum of a convergent power series with center z_0,

$$f(z) = \sum_{k=0}^{\infty} a_k\,(z - z_0)^k, \qquad |z - z_0| < r. \tag{3.28}$$

Note that, according to this definition, the analyticity of f at z_0 implies that z_0 is an inner point of the domain set D.

Uniqueness of Representation. In principle, an analytic function could have different representations (3.28) as power series at z_0. In order to prove that this cannot happen, we investigate to which extent the coefficients of a power series are determined by the values of its sum.

Theorem 3.2.6 (Uniqueness Principle, Identity Theorem)**.** *Let f and g be the sums of two power series with center z_0, and assume that both converge in an open disk $D := \{z \in \mathbb{C} : |z - z_0| < r\}$,*

$$f(z) = \sum_{k=0}^{\infty} a_k\,(z - z_0)^k, \qquad g(z) = \sum_{k=0}^{\infty} b_k\,(z - z_0)^k. \tag{3.29}$$

If there exists a sequence $(z_n) \subset D \setminus \{z_0\}$ such that $z_n \to z_0$ and $f(z_n) = g(z_n)$ for all $n \in \mathbb{N}$, then $a_k = b_k$ for all $k \in \mathbb{N}$ and $f(z) = g(z)$ for all $z \in D$.

Proof. By Theorem 3.2.3 the functions f and g are continuous at z_0, and hence

$$a_0 = f(z_0) = \lim_{n\to\infty} f(z_n) = \lim_{n\to\infty} g(z_n) = g(z_0) = b_0.$$

Using the arithmetic rules for convergent sequences, we obtain the representations

$$f_1(z) := \frac{f(z) - a_0}{z - z_0} = \sum_{k=0}^{\infty} a_{k+1}\,(z - z_0)^k, \quad g_1(z) := \frac{g(z) - b_0}{z - z_0} = \sum_{k=0}^{\infty} b_{k+1}\,(z - z_0)^k,$$

for all $z \in D \setminus \{z_0\}$. Because of $a_0 = b_0$ we have $f_1(z_n) = g_1(z_n)$ for all $n \in \mathbb{N}$, which implies $a_1 = b_1$, as has just been shown. Proceeding inductively, we get $a_k = b_k$ for all k, and finally $f(z) = g(z)$ for all $z \in D$. ☐

The fact that the sum of a convergent power series determines the coefficients of that series uniquely is of fundamental importance for the technique of comparing coefficients which we shall study soon.

Taylor Coefficients. The coefficients a_k of the power series (3.28) representing a function f analytic at z_0 are referred to as the *Taylor coefficients* of f at z_0. The series (3.28) itself is said to be the *Taylor series* of f with center z_0. So we can say that a function f is analytic at z_0 if it admits a convergent Taylor series expansion with center z_0.

Theoretically, the Taylor coefficients a_k can be determined by the limit procedures in the proof of Theorem 3.2.6, but this is inappropriate for practical purposes. In Section 4.1 and Section 4.2 we shall derive several explicit formulas (see (4.5), (4.22), (4.45)) to do this in an easier and more elegant manner.

Knowing that the Taylor coefficients are uniquely determined, we may ask how operations with functions analytic at a point z_0 are reflected in these coefficients. For example, adding up the power series (3.29) term by term, like for any convergent series, we obtain convergent power series for $f + g$ and $f - g$,

$$f(z) \pm g(z) = \sum_{k=0}^{\infty} (a_k \pm b_k)\,(z - z_0)^k, \quad z \in D, \tag{3.30}$$

so that $f+g$ and $f-g$ are analytic at z_0. Establishing analogous statements for products, quotients and compositions will occupy us for a while, so a relaxing interlude may be appropriate. We also use the opportunity to introduce two new functions and to demonstrate how phase portraits can help us to find functional relations.

Phase Portrait Experiments. Before we begin, let us first outline the philosophy behind these experiments. Recall that a function f may be considered as a black box, which takes an input and produces an output. Usually the input is an argument z and the output is the corresponding value $f(z)$. In our experiments we take a somewhat different perspective: the input is a domain of the z-plane, usually a square Q, and the output is the phase portrait $P_f(Q)$ of the restriction of f to that domain. The goal is to derive as much information as possible about the function f from the phase portrait, *without having to take recourse to the values of the function.*

In other words, we think of a function as an unknown object, which manifests itself in a phase portrait. Working in a laboratory for exploring functions, phase portraits are *measuring devices* which allow us to extract information. Like in physics, just one experiment will usually not suffice to understand all aspects of the object of investigation, so we shall make a series of experiments with modified settings. For example, we may take a known "test" function g and consider the phase portraits of the sum $f+g$, the product fg, or the composition $g \circ f$. Like in the natural sciences, the *design* of these experiments is crucial for their success. An experimenter needs training and experience to know or to guess which instruments could be useful in certain situations.

Note that there is another parallel to physics: the accuracy of measurements is limited, phase portraits usually depict only part of the function with a finite resolution. So we have to make sure whether the effects which we observe, or believe we observe, "really" happen. Fortunately, mathematics provides us with a reliable tool to crosscheck observations: the proof. So we shall be using phase portraits as a source of ideas and an aid to geometric reasoning, but eventually all conjectures must be verified.

Exploring the Sine and Cosine Functions. The *complex sine* and *cosine* functions are defined by evaluating the Taylor series of the corresponding real functions for complex arguments z,

$$\sin z = \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k+1)\,!}\, z^{2k+1} = z - \frac{z^3}{3\,!} + \frac{z^5}{5\,!} \mp \ldots + (-1)^k \, \frac{z^{2k+1}}{(2k+1)\,!} + \ldots \tag{3.31}$$

$$\cos z = \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k)\,!}\, z^{2k} = 1 - \frac{z^2}{2\,!} + \frac{z^4}{4\,!} \mp \ldots + (-1)^k \, \frac{z^{2k}}{(2k)\,!} + \ldots . \tag{3.32}$$

It is easy to see that the series converge for all $z \in \mathbb{C}$. Figure 3.15 shows the enhanced phase portraits of both functions in the square $|\mathrm{Re}\, z| < 4$, $|\mathrm{Im}\, z| < 4$.

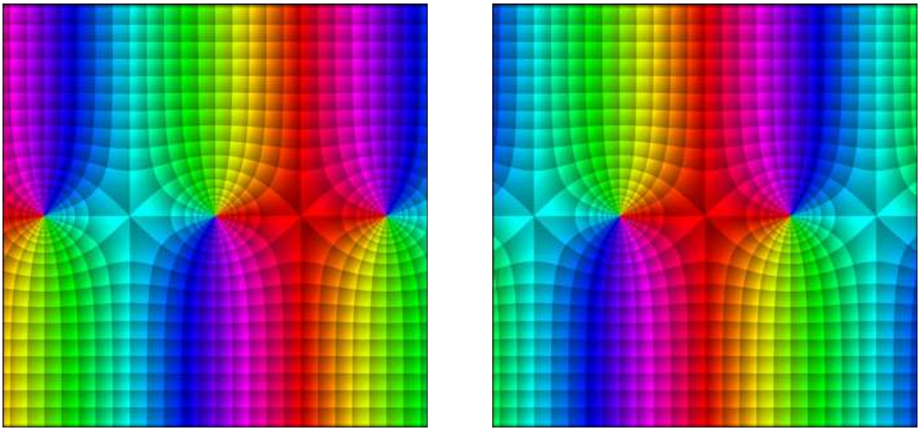


Figure 3.15: Enhanced phase portraits of the complex sine and cosine function

The most striking observation is that, apart from a translation, the phase portraits seem to be identical, though the defining series of the functions are totally different. Another remarkable fact is their invariance with respect to a horizontal shift, which becomes visible more clearly when the domain depicted is enlarged, as we did for the cosine function in Figure 3.16 (left). Furthermore we notice that at some distance off the real axis the phase portrait is composed of parallel vertical stripes. And here is a first rule of thumb in our training on phase portraits:

Whenever the phase portrait shows (almost) parallel stripes, the exponential function might be involved.

Indeed the stripes in the phase portraits are crucial for understanding the sine and the cosine functions. Since the deductions are slightly simpler for the cosine function, we work with it and leave the sine function as an exercise for the reader.

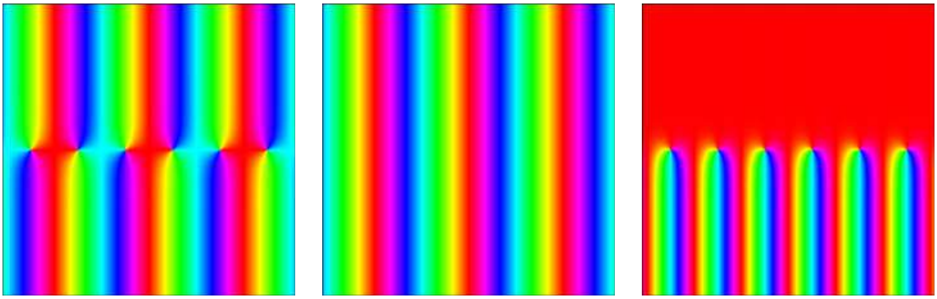


Figure 3.16: Phase portraits of $f(z) = \cos z$, $g(z) = \mathrm{e}^{-\mathrm{i}\, z}$, and $f(z)/g(z)$

Because the stripes in the phase portrait of the cosine function are vertical, we transform the exponential function appropriately, pre-composing it with the mapping $z \mapsto -\mathrm{i}\, z$, which rotates the phase portrait counterclockwise (!) by an angle of $\pi/2$. The result is shown in the middle of Figure 3.16. Now both images almost coincide in the upper region. To continue, we apply a second rule:

If the phase portraits of two functions f and g look "similar", then the phase portrait of their quotient f/g should be monochromatic.

The result depicted in Figure 3.16 (right) shows that we have been successful. In the upper domain the phase portrait of f/g is almost isochromatic. Since the red color corresponds to positive real values, we conclude that the quotient f/g is close to a positive function in that domain. This does not say too much about the actual values of the quotient, but at the moment we are just experimenting, and so we may apply another heuristic principle, known as

Ockham's razor: *Simpler explanations are generally better than complicated ones.*

The simplest explanation for our findings could be that f is almost a *constant* positive multiple of g, $f \approx c\, g$. Assuming, that this is really so, how can we determine the value of the positive constant c from observations made from appropriate phase portraits ?

The key to answering this question lies in *tuning experiments*, where we consider the phase portraits of $f - c\, g$ for different positive values of c. If c is close to the correct value, then $f - c\, g$ should be close to the zero function, and we can hope that even small changes in the value of c will have a strong effect on the phase portrait.

Design of tuning experiments: if f/g is almost a positive constant, consider differences $f - cg$ for various positive values of c. At the correct value of c the phase portrait changes abruptly.

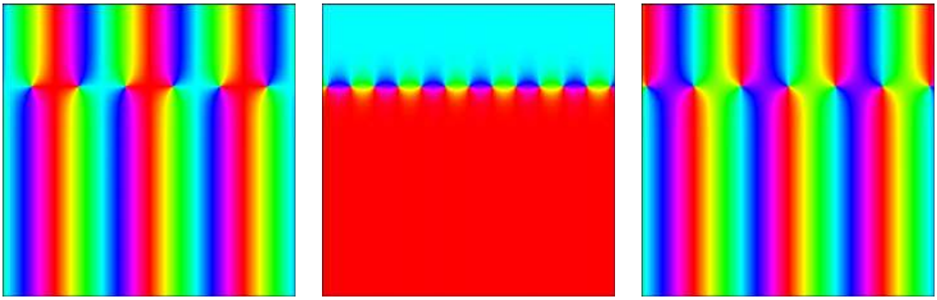


Figure 3.17: Tuning experiments for the constant c in $f - c\, g$

The two outer windows of Figure 3.17 show the phase portraits of the functions $\cos z - c\, \exp(\mathrm{i}\, z)$ for $c_1 = 0.4999$ and $c_2 = 0.5001$, respectively. In their lower parts both pictures are almost identical, while they are quite different in the upper parts (note the phase shift).

As was already mentioned, the phase portraits of two functions can be conveniently compared by considering the phase portrait of their quotient. In our case this is the function

$$h(z) := \frac{f(z) - c_1\, g(z)}{f(z) - c_2\, g(z)} = \frac{\cos z - c_1\, \exp(\mathrm{i}\, z)}{\cos z - c_2\, \exp(\mathrm{i}\, z)},$$

depicted in the middle of Figure 3.17. In order to interpret this picture, we remark that the values $f(z) - c_1\, g(z)$ and $f(z) - c_2\, g(z)$ differ only by a small amount if c_1 and c_2 are close. So, typically both values also have almost the same phase – except in the case where both are close to zero. Consequently, the phase of the quotient $h(z)$ can significantly differ from 1 only if $f(z) \approx c_1\, g(z) \approx c_2\, g(z)$, which must happen in those regions of the phase portrait of h which are *not* colored red.

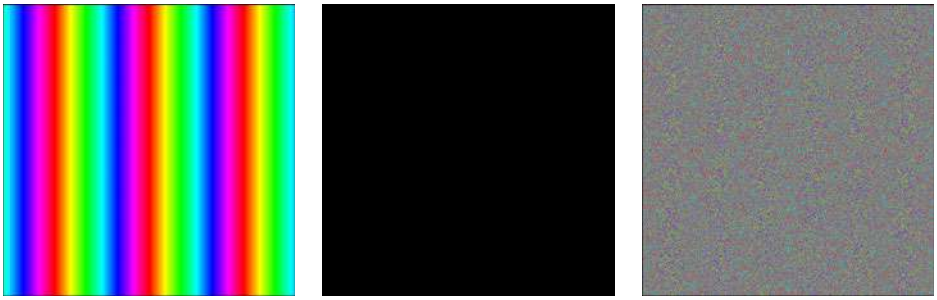


Figure 3.18: Phase portraits of $\cos z - \mathrm{e}^{\mathrm{i}\, z}/2$, $h(z)$, and a numerical zero function.

The results of these experiments suggest that the best constant c might be $1/2$. The phase portrait of $\cos z - (1/2)\, \exp(\mathrm{i}\, z)$ is shown in Figure 3.18 (left). It resembles the phase portrait of $\exp(-\mathrm{i}\, z)$, and tuning experiments for the functions

$$\cos z - \frac{1}{2}\, \mathrm{e}^{\mathrm{i}\, z} - c\, \mathrm{e}^{-\mathrm{i}\, z}$$

lead to the conjecture that $c = 1/2$ is again the best choice for c. The window in the middle depicts the phase portrait of the function

$$h(z) := \cos z - \frac{1}{2}\, \mathrm{e}^{\mathrm{i}\, z} - \frac{1}{2}\, \mathrm{e}^{-\mathrm{i}\, z}.$$

Since this window is completely black, h is the zero function. In a real experiment the function h is usually computed numerically with limited accuracy. Because the numerical errors cause fluctuations, the computed values $h(z)$ are not exactly all zero. Hence we would then see something like what we have in the picture on the right in Figure 3.18.

Finally, the conjecture that h is the zero function can easily be verified by adding the power series for the functions $\exp(\mathrm{i}\,z)$ and $\exp(-\mathrm{i}\,z)$ according to (3.30),

$$\begin{aligned}
\mathrm{e}^{\mathrm{i}z}+\mathrm{e}^{-\mathrm{i}z} &= \sum_{k=0}^{\infty}\frac{(\mathrm{i}\,z)^k}{k\,!}+\sum_{k=0}^{\infty}\frac{(-\mathrm{i}\,z)^k}{k\,!}\\
&= \sum_{k=0}^{\infty}\left(\frac{z^{4k}}{(4k)\,!}+\mathrm{i}\,\frac{z^{4k+1}}{(4k+1)\,!}-\frac{z^{4k+2}}{(4k+2)\,!}-\mathrm{i}\,\frac{z^{4k+3}}{(4k+3)\,!}\right)\\
&\qquad+\sum_{k=0}^{\infty}\left(\frac{z^{4k}}{(4k)\,!}-\mathrm{i}\,\frac{z^{4k+1}}{(4k+1)\,!}-\frac{z^{4k+2}}{(4k+2)\,!}+\mathrm{i}\,\frac{z^{4k+3}}{(4k+3)\,!}\right)\\
&= 2\sum_{k=0}^{\infty}\left(\frac{z^{4k}}{(4k)\,!}-\frac{z^{4k+2}}{(4k+2)\,!}\right)=2\sum_{k=0}^{\infty}(-1)^k\,\frac{z^{2k}}{(2k)\,!}=2\,\cos z.
\end{aligned}$$

A similar calculation yields a representation of the sine function by exponential functions. In summary, we have

$$\cos z=\frac{1}{2}\left(\mathrm{e}^{\mathrm{i}z}+\mathrm{e}^{-\mathrm{i}z}\right),\quad \sin z=\frac{1}{2\mathrm{i}}\left(\mathrm{e}^{\mathrm{i}z}-\mathrm{e}^{-\mathrm{i}z}\right),\quad z\in\mathbb{C}. \tag{3.33}$$

Multiplying the second equation by i and adding the result to the first equality, we get

$$\mathrm{e}^{\mathrm{i}\,z}=\cos z+\mathrm{i}\,\sin z,\qquad z\in\mathbb{C}. \tag{3.34}$$

For real z, we obtain a representation of the complex exponential function on the imaginary axis by the *real* sine and cosine functions. In particular, e^z is *unimodular* on the imaginary axis,

$$\left|\mathrm{e}^{\mathrm{i}\,y}\right|=1,\quad y\in\mathbb{R}.$$

Note that (3.34) allows us to rewrite the polar representation of a complex number as

$$z=r\,(\cos\varphi+\mathrm{i}\sin\varphi)=r\,\mathrm{e}^{\mathrm{i}\varphi}.$$

For $\varphi=\pi$ we obtain the miraculous formula

$$\mathrm{e}^{\mathrm{i}\pi}+1=0,$$

which is the unbeatable winner of a beauty contest for mathematical formulas. Involving the three basic arithmetic operations: addition, multiplication, exponentiation and the equality sign, it establishes a relation between the five fundamental constants 0, 1, e, π and i. Most remarkably, the first four numbers seem to be unrelated in real analysis, only the imaginary unit is the key that reveals their intimate connection.

We mention that all results derived experimentally could have been obtained just by adding the power series of $\cos z$ and $\mathrm{i}\,\sin z$, which yields (3.34). We have deliberately chosen to use phase portraits in order to play around a little with these objects and also to demonstrate how properties of an unknown complex function can be *seen* in appropriately designed experiments.

Operations Involving Analytic Function. After this interlude let us continue with our main theme. The following theorem summarizes qualitative statements on the analyticity of sums, products, quotients and compositions.

Theorem 3.2.7. *If f and g are analytic at z_0, then $f+g$, $f-g$ and $f\,g$ are analytic at z_0. If, moreover, $g(z_0)\neq 0$, then f/g is analytic at z_0. If f is analytic at z_0 and g is analytic at $w_0 := f(z_0)$, then $g\circ f$ is analytic at z_0.*

Note that the composition $g\circ f$ need not exist on the domain of f, but just in a sufficiently small neighborhood of z_0.

The analyticity of $f+g$ and $f-g$ has already been verified. The proofs of the remaining assertions are more demanding. Quantitative versions of these statements will be obtained in Theorems 3.2.8, 3.2.9 and 3.2.11.

The Cauchy Product. The next result is a more sophisticated statement about the analyticity of a product $f\,g$, which includes an algorithm for computing the Taylor coefficients of $f\,g$ from the coefficients of the factors f and g.

Theorem 3.2.8 (Cauchy Product)**.** *Assume that the power series (3.29) for f and g converge in an open disk D, and let*

$$c_k := \sum_{j=0}^{k} a_j b_{k-j}.$$

Then the power series $\sum c_k\,(z-z_0)^k$ converges in D to the product $f(z)\,g(z)$,

$$f(z)\,g(z) = \sum_{k=0}^{\infty} c_k\,(z-z_0)^k, \qquad z\in D. \tag{3.35}$$

The series (3.35) is said to be the *Cauchy product* of the series (3.29).

Proof. Let f_n, g_n, and p_n be the partial sums of the series in (3.29) and (3.35), respectively. Then a rearrangement of the finite sums yields

$$\begin{aligned} f_n(z)\,g_n(z) &= \Big(\sum_{j=0}^{n} a_j\,(z-z_0)^j\Big)\Big(\sum_{i=0}^{n} b_i\,(z-z_0)^i\Big) = \sum_{i=0}^{n}\sum_{j=0}^{n} a_j b_i\,(z-z_0)^{i+j} \\ &= \sum_{k=0}^{n}\sum_{j=0}^{k} a_j b_{k-j}\,(z-z_0)^k + \sum_{k=n+1}^{2n}\sum_{j=k-n}^{n} a_j b_{k-j}\,(z-z_0)^k \\ &= p_n(z) + \sum_{k=n+1}^{2n}\sum_{j=k-n}^{n} a_j b_{k-j}\,(z-z_0)^k. \end{aligned}$$

Fix $z\in D$. Since D is an open disk, there exists r such that $r>|z-z_0|$ and $z_0+r\in D$. Then Lemma 3.2.2 tells us that $|a_k|\le c\,r^{-k}$ and $|b_k|\le c\,r^{-k}$ for some constant c and all k. Setting $q := |z - z_0|/r < 1$, we use the triangle inequality to estimate

$$\begin{aligned}|f_n(z)\,g_n(z) - p_n(z)| &\leq \sum_{k=n+1}^{2n} \sum_{j=k-n}^{n} |a_j|\,|b_{k-j}|\,|z-z_0|^k \\ &\leq \sum_{k=n+1}^{2n} \sum_{j=k-n}^{n} c^2\, r^{-k}|z-z_0|^k \leq \sum_{k=n+1}^{2n} (2n-k+1)\, c^2 q^k \leq n^2\, c^2\, q^{n+1}.\end{aligned}$$

Since the right-hand side tends to zero as $n \to \infty$, the assertion follows. □

Example 3.2.4 (Exponential Function). Let $x, y \in \mathbb{C}$. Computing the Cauchy product of the Taylor series of $\exp x$ and $\exp y$, we obtain the *addition theorem* of the exponential function,

$$\begin{aligned}\mathrm{e}^x \cdot \mathrm{e}^y = \Big(\sum_{k=0}^{\infty} \frac{x^k}{k!}\Big) \cdot \Big(\sum_{j=0}^{\infty} \frac{y^j}{j!}\Big) &= \sum_{k=0}^{\infty} \Big(\sum_{j=0}^{k} \frac{x^j}{j!}\, \frac{y^{k-j}}{(k-j)!}\Big) \\ = \sum_{k=0}^{\infty} \sum_{j=0}^{k} \frac{1}{k!} \binom{k}{j}\, x^j y^{k-j} &= \sum_{k=0}^{\infty} \frac{1}{k!} (x+y)^k = \mathrm{e}^{x+y}.\end{aligned}$$

When this identity is applied to $z = x + \mathrm{i}\,y$ with $x, y \in \mathbb{R}$, it yields a representation of the complex exponential function by familiar real functions,

$$\mathrm{e}^{x+\mathrm{i}\,y} = \mathrm{e}^x\,(\cos y + \mathrm{i}\,\sin y),$$

which implies that for all $z \in \mathbb{C}$

$$\big|\mathrm{e}^z\big| = \mathrm{e}^{\operatorname{Re} z}, \quad \arg\, \mathrm{e}^z = \operatorname{Im} z, \quad \mathrm{e}^{z+2\pi\mathrm{i}} = \mathrm{e}^z. \tag{3.36}$$

In particular, the exponential function has no zeros and is *periodic* with the purely imaginary period $2\pi\mathrm{i}$. The second and the third relations in (3.36) are nicely reflected in the phase portrait of the exponential function, see Figure 3.13 (right).

Using the addition theorem and (3.33) it is now also easy to verify that the zeros of the sine and the cosine functions are precisely $k\pi$ and $\pi/2 + k\pi$ for integral values of k, respectively.

Reciprocal Functions. In order to obtain the assertion of Theorem 3.2.7 about the division of analytic functions it suffices to verify that the *reciprocal* $1/f$ of f is analytic and then to apply Theorem 3.2.8.

Theorem 3.2.9. *If f is analytic at z_0 and $f(z_0) \neq 0$, then $1/f$ is analytic at z_0. The Taylor coefficients b_k of $1/f$ at z_0 can be computed recursively from the Taylor coefficients a_k of f by $b_0 := 1/a_0$ and*

$$b_k := -\frac{1}{a_0}\,\big(a_1 b_{k-1} + a_2 b_{k-2} + \ldots + a_k b_0\big), \quad k = 1, 2, \ldots. \tag{3.37}$$

Proof. 1. In the first step we *assume* that the function $1/f$ is analytic at z_0. Then its Taylor series

$$\frac{1}{f(z)} = b_0 + b_1\,(z-z_0) + b_2\,(z-z_0)^2 + \ldots + b_k\,(z-z_0)^k + \ldots \tag{3.38}$$

converges in a neighborhood of z_0 and its Cauchy product with the Taylor series of f is the constant function 1. The latter is equivalent to the infinite system of equations

$$\begin{aligned} a_0 b_0 &= 1 \\ a_0 b_1 + a_1 b_0 &= 0 \\ a_0 b_2 + a_1 b_1 + a_2 b_0 &= 0 \\ &\vdots \end{aligned}$$

Since $a_0 \neq 0$, this triangular system can be solved with respect to the coefficients b_k, which yields the recursion (3.37).

2. It remains to prove that the series (3.38), with coefficients b_k given by the recursion (3.37), indeed has a positive radius of convergence.

By Cauchy's estimate (3.22) in Lemma 3.2.2, there are positive numbers c and r such that $|a_n| \le c r^{-n}$ for all $n \in \mathbb{N}$. We set $q := 1 + c/|a_0|$ and show that

$$|b_n| \le \frac{c}{|a_0|^2}\,\frac{q^{n-1}}{r^n}, \qquad n = 1, 2, \ldots. \tag{3.39}$$

For $n = 1$ we have $b_1 = -a_1/a_0^2$ and $|a_1| \le c/r$, so that indeed

$$|b_1| = \frac{a_1}{a_0^2} \le \frac{c}{|a_0|^2}\,\frac{1}{r}.$$

Now we assume that (3.39) holds for all $n = 1, 2, \ldots, k-1$ and consider the case where $n = k$. Using $|b_0| = 1/|a_0|$, the recursive definition of b_k, and the triangle inequality, we estimate

$$\begin{aligned} |b_k| &\le \frac{1}{|a_0|}\left(|a_k b_0| + \sum_{j=1}^{k-1} |a_{k-j}|\,|b_j|\right) \\ &\le \frac{1}{|a_0|}\left(|a_k b_0| + \sum_{j=1}^{k-1} \frac{c}{r^{k-j}}\,\frac{c}{r^j |a_0|^2}\,q^{j-1}\right) \\ &\le \frac{c}{r^k\,|a_0|^2}\left(1 + \frac{c}{|a_0|}\sum_{j=0}^{k-2} q^j\right) \\ &= \frac{c}{r^k\,|a_0|^2}\left(1 + \frac{c}{|a_0|}\,\frac{q^{k-1}-1}{q-1}\right) = \frac{c}{r^k\,|a_0|^2}\,q^{k-1}, \end{aligned}$$

which gives (3.39) for $n = k$ and thus for all n. Consequently, by Lemma 3.2.2, the power series (3.38) has radius of convergence not less than r/q. □

Example 3.2.5 (Bernoulli Numbers). Let the function f be defined on the complex plane by $f(z) := (\mathrm{e}^z - 1)/z$ if $z \neq 0$ and $f(0) := 1$. Representing e^z by its Taylor series, we obtain the power series

$$f(z) := \sum_{k=0}^{\infty} a_k\, z^k = \sum_{k=0}^{\infty} \frac{z^k}{(k+1)\,!}$$

which converges in the entire complex plane and attains the correct value $f(0) = 1$ at $z = 0$. Since $f(0) \neq 0$, the reciprocal function $1/f$ is also analytic at $z_0 = 0$. Figure 3.19 shows the enhanced phase portraits of the functions f and $1/f$ in the square $|\operatorname{Re} z| < 15$, $|\operatorname{Im} z| < 15$.

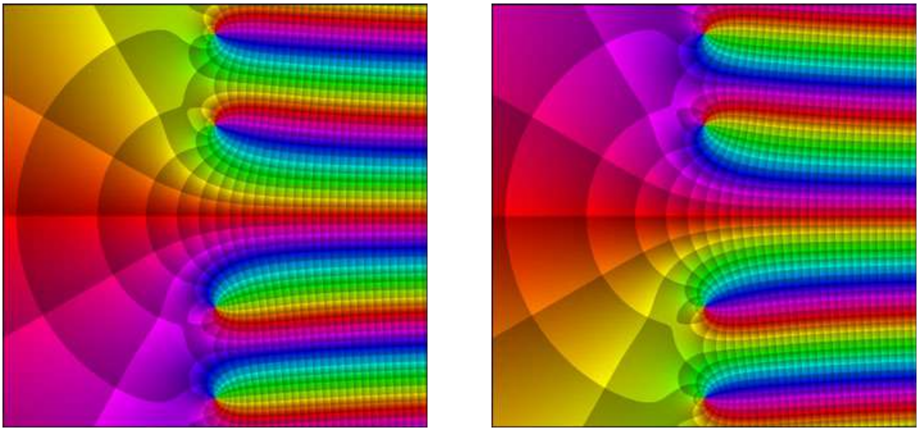


Figure 3.19: Enhanced phase portraits of $f(z) = (\mathrm{e}^z - 1)/z$ and $1/f(z)$

Writing the Taylor series of $g := 1/f$ in the form

$$g(z) = \frac{z}{\mathrm{e}^z - 1} = \sum_{k=0}^{\infty} b_k\, z^k = \sum_{k=0}^{\infty} \frac{B_k}{k\,!}\, z^k, \tag{3.40}$$

the numbers B_k are determined by the equations $B_0 = b_0 = 1/a_0 = 1$ and

$$0 = \sum_{j=0}^{k} a_{k-j}\, b_j = \sum_{j=0}^{k} \frac{1}{(k-j+1)!}\,\frac{B_j}{j\,!} = \frac{1}{(k+1)\,!}\sum_{j=0}^{k} \binom{k+1}{j} B_j,$$

for $k = 1, 2, \ldots$. Solving this system recursively, we get

$$B_k = -\frac{1}{k+1}\sum_{j=0}^{k-1} \binom{k+1}{j} B_j, \quad k = 1, 2, \ldots.$$

The numbers B_k are called the *Bernoulli numbers*. For n odd, all B_n are zero, except for B_1 which equals $-1/2$. The first Bernoulli numbers for n even are

$$B_2 = \frac{1}{6},\ B_4 = -\frac{1}{30},\ B_6 = \frac{1}{42},\ B_8 = -\frac{1}{30},\ B_{10} = \frac{5}{66},\ B_{12} = -\frac{691}{2730}.$$

Note that the series (3.40) converges for $|z| < 2\pi$. We shall not prove this here, since it will follow immediately from a more general result which we shall derive later (Theorem 4.2.24).

Example 3.2.6 (Tangent and Cotangent). The complex *tangent* and *cotangent* function are defined as quotients of the sine and cosine function,

$$\tan z := \frac{\sin z}{\cos z}, \qquad \cot z := \frac{\cos z}{\sin z}. \tag{3.41}$$

Both functions are π-periodic. The zeros of $\tan z$ and $\cot z$ are located at the points $k\pi$ and $\pi/2 + k\pi$ with $k \in \mathbb{Z}$, respectively. Of course the poles of one function are the zeros of the other. Figure 3.20 shows the enhanced phase portraits of these functions in the square $|\mathrm{Re}\, z| < 4$, $|\mathrm{Im}\, z| < 4$.

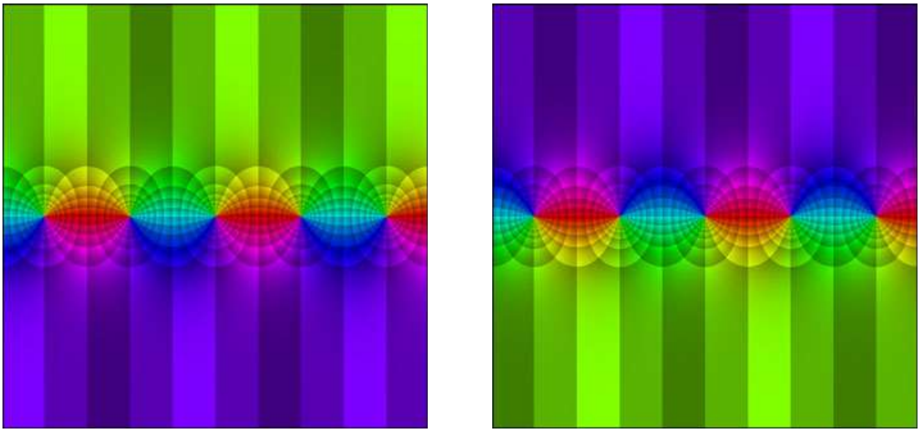


Figure 3.20: Enhanced phase portraits of the tangent and the cotangent function

Since the sine and the cosine functions are analytic at $z_0 = 0$, and $\cos 0 \neq 0$, the tangent function is also analytic at this point. Instead of computing the Taylor coefficients as solutions of the linear system (3.37) we derive the Taylor series directly from the preceding example.

A straightforward calculation shows that for all $z \in \mathbb{C}$ with $\mathrm{e}^{4\mathrm{i}z} \neq 1$

$$\tan z = \frac{1}{\mathrm{i}} \frac{\mathrm{e}^{\mathrm{i}z} - \mathrm{e}^{-\mathrm{i}z}}{\mathrm{e}^{\mathrm{i}z} + \mathrm{e}^{-\mathrm{i}z}} = \frac{1}{\mathrm{i}} \left(1 - \frac{2}{\mathrm{e}^{2\mathrm{i}z} - 1} + \frac{4}{\mathrm{e}^{4\mathrm{i}z} - 1} \right).$$

Consequently, defining the function g by $g(z) := z/(\mathrm{e}^z - 1)$, we have

$$z \tan z = \frac{z}{\mathrm{i}} + g(2\mathrm{i}z) - g(4\mathrm{i}z).$$

Representing g by its Taylor series (3.40), we obtain

$$z \tan z = \frac{z}{\mathrm{i}} + \sum_{k=0}^{\infty} \frac{B_k}{k\,!} \left((2\mathrm{i})^k - (4\mathrm{i})^k\right) z^k.$$

The summand with $k = 0$ vanishes and, thanks to $B_1 = -1/2$, the summand with $k = 1$ just compensates the additional term z/i. Taking into account that the Bernoulli numbers B_{2k+1} with $k > 1$ are zero, we need only extend the sum over even k and get

$$z \tan z = \sum_{k=1}^{\infty} \frac{B_{2k}}{(2k)\,!} (-1)^k\, 2^{2k} \left(1 - 2^{2k}\right) z^{2k}.$$

The construction guarantees that this equality holds for all z with $0 < |4z| < 2\pi$ (see the remark at the end of the preceding example), but obviously it is also satisfied for $z = 0$. Finally, dividing by z, we arrive at

$$\tan z = \sum_{k=1}^{\infty} (-1)^{k+1} \frac{B_{2k}}{(2k)\,!}\, 4^k (4^k - 1)\, z^{2k-1}. \tag{3.42}$$

This series has radius of convergence $\pi/2$.

Since $\cot 0 = \infty$, the *cotangent function* cannot be represented as a convergent power series at $z_0 = 0$. As a substitute we consider the function f defined by $f(z) := z \cot z$ for $z \neq 0$ and $f(0) := 1$. This function is analytic at $z_0 = 0$ and has the Taylor series

$$z \cot z = \sum_{k=0}^{\infty} (-1)^k\, B_{2k}\, \frac{4^k}{(2k)\,!}\, z^{2k},$$

with radius of convergence π.

Remark 3.2.10 (**Power Series and Matrix Algebra**). Arithmetic operations with power series can be reduced to matrix algebra by associating with the series $\sum a_k\,(z - z_0)^k$ the infinite upper triangular matrix

$$A = \begin{bmatrix} a_0 & a_1 & a_2 & a_3 & \dots \\ 0 & a_0 & a_1 & a_2 & \dots \\ 0 & 0 & a_0 & a_1 & \dots \\ 0 & 0 & 0 & a_0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}.$$

Matrices of this type are called *semi-circulant*. The sum $S = A + B$ and the product $P = AB$ of semi-circulant matrices A and B are defined as the semi-circulant matrices with entries

$$s_{nk} := a_{nk} + b_{nk}, \qquad p_{nk} := \sum_{j=n}^{k} a_{nj}\, b_{jk},$$

respectively, where $p_{nk} := 0$ if the sum is void (which happens if $n > k$). Note that sum and product operations of semi-circulant matrices are commutative.

If A and B are the semi-circulant matrices associated with $\sum a_k\,(z - z_0)^k$ and $\sum b_k\,(z - z_0)^k$, respectively, then the sum and the product of these series are associated to the matrices $A + B$ and AB.

A (semi-circulant) matrix A is called *invertible* if there exists a (semi-circulant) matrix B such that $AB = I$ (which implies that $BA = I$). Here I is the *identity matrix*, which has entries 1 on the main diagonal and 0 elsewhere. If $AB = I$, then B is uniquely defined, it is said to be the *inverse* of A, and we denote it by A^{-1}.

It is easily seen that A is invertible if and only if $a_0 \neq 0$. Then *Cramer's rule* delivers an explicit formula for the entries b_k of the inverse A^{-1}, which is known as *Wronski's formula*, namely $b_0 = 1/a_0$ and

$$b_k = \frac{(-1)^k}{a_0^{n+1}} \det \begin{bmatrix} a_1 & a_2 & a_3 & \dots & a_{n-1} & a_n \\ a_0 & a_1 & a_2 & \dots & a_{n-2} & a_{n-1} \\ 0 & a_0 & a_1 & \dots & a_{n-3} & a_{n-2} \\ \vdots & \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & a_0 & a_1 \end{bmatrix}, \qquad n = 1, 2, \dots. \tag{3.43}$$

In particular, this formula gives an explicit representation of the coefficients b_k of the reciprocal power series of $\sum a_k (z - z_0)^k$.

Composition of Power Series. The final step in proving Theorem 3.2.7 is concerned with the composition $g \circ f$ of functions given by power series. In order to ensure that the composition makes sense at least locally, we assume that f is analytic at z_0, while g is supposed to be analytic at the image point $w_0 := f(z_0)$. Then, by continuity, f maps a neighborhood of z_0 into the disk of convergence of g. Our goal is to find a convergent power series for $g \circ f$ from the given power series of f and g. The approach is straightforward: we assume that

$$f(z) = \sum_{k=0}^{\infty} a_k\,(z - z_0)^k, \qquad g(w) = \sum_{k=0}^{\infty} b_k\,(w - w_0)^k, \tag{3.44}$$

substitute $w - w_0 = \sum_{n=1}^{\infty} a_n (z - z_0)^n$ in the series for g, rearrange the double sum according to the powers of $z - z_0$ (note that for every k the number of summands involving $(z - z_0)^k$ is finite), and show that the resulting series converges to $g \circ f$ in a neighborhood of z_0. The details will be worked out next.

For $n = 1, 2, \ldots$, the nth power $(f - a_0)^n$ is analytic at z_0 and the n leading terms of its Taylor series at z_0 vanish. Denoting by a_{nk} the Taylor coefficients of this function, we have

$$\big(f(z) - a_0\big)^n = \sum_{k=1}^{\infty} a_{nk}\,(z - z_0)^k = \sum_{k=n}^{\infty} a_{nk}\,(z - z_0)^k, \quad n = 1, 2, \ldots, \tag{3.45}$$

in some neighborhood of z_0. Substituting the term $w - w_0$ in the power series of $g - b_0$ by the power series of $f - a_0$ (recall that $a_0 = f(z_0) = w_0$), we obtain formally

$$\sum_{n=1}^{\infty} b_n\,(w{-}w_0)^n = \sum_{n=1}^{\infty} b_n \left(\sum_{k=n}^{\infty} a_{nk}\,(z{-}z_0)^k\right) = \sum_{k=1}^{\infty} \left(\sum_{n=1}^{k} b_n a_{nk}\right)(z{-}z_0)^k. \tag{3.46}$$

Before we justify that changing the order of summation is possible, we state the result.

Theorem 3.2.11. *If f is analytic at z_0 and g is analytic at $w_0 := f(z_0)$, then $g \circ f$ is analytic at z_0. Let f, g, and $(f - a_0)^n$ be represented by the series (3.44) and (3.45), respectively. Then the Taylor coefficients c_k of $g \circ f$ at z_0 are given by*

$$c_0 = b_0, \qquad c_k = \sum_{n=1}^{k} b_n\, a_{nk}, \quad k = 1, 2, \ldots. \tag{3.47}$$

Proof. We choose a positive number r which is smaller than the radii of convergence of the Taylor series of f and g and consider $z \in \mathbb{C}$ with $|z - z_0| \le r$. If r is sufficiently small, then, by continuity of f, $w := f(z)$ lies in the disk of convergence of the Taylor series of g, and the left-hand side of (3.46) is equal to $(g \circ f)(z)$.

The first equality of (3.46) follows directly from the definition of the Taylor coefficients a_{nk} in (3.45).

In order to justify the second equality of (3.46) it suffices to show that one of the double series converges absolutely if $|z - z_0|$ is sufficiently small, for example

$$\sum_{n=1}^{\infty} |b_n| \sum_{k=n}^{\infty} |a_{nk}|\,|z - z_0|^k < \infty, \qquad |z - z_0| < \delta. \tag{3.48}$$

Let

$$c := \max_{|z - z_0| \le r} |f(z) - a_0|, \qquad d := \max_{|w - w_0| \le r} |g(w) - b_0|.$$

Since the a_{nk} are the Taylor coefficients of $(f - a_0)^n$, and

$$\max_{|z - z_0| \le r} \left|\big(f(z) - a_0\big)^n\right| = c^n,$$

we have by Cauchy's estimate in Lemma 3.2.4,

$$|a_{nk}| \le c^n\, r^{-k}, \qquad |b_n| \le d\, r^{-n}. \tag{3.49}$$

Consequently, if δ is chosen so small that $\delta < \min(r, r^2/c)$, then for all $M, N \in \mathbb{N}$,

$$\sum_{n=1}^{N} |b_n| \sum_{k=n}^{M} |a_{nk}|\,|z-z_0|^k \le \sum_{n=1}^{N} \frac{d}{r^n} \sum_{k=n}^{M} \frac{c^n}{r^k}\,\delta^k \le \frac{dr}{r-\delta} \sum_{n=1}^{N} \left(\frac{c\delta}{r^2}\right)^n \le \frac{dr}{r-\delta}\,\frac{r^2}{r^2-c\delta},$$

and the sequence (3.48) converges by Cauchy's criterion. So, for all z sufficiently close to z_0, the left-hand side of (3.46) is equal to $(f \circ g)(z) - a_0$, and comparing the right-hand side with the definition (3.47) of the c_k we obtain the desired result. ☐

3.3 Introduction to Analytic Functions

Recall that a function $f : D \to \widehat{\mathbb{C}}$ is said to be analytic at a point $z_0 \in D$, if it admits a power series expansion at z_0, i.e., there exists an open disk D_0 with center at z_0 such that D_0 is contained in D and

$$f(z) = a_0 + a_1\,(z-z_0) + \ldots + a_k\,(z-z_0)^k + \ldots \qquad z \in D_0. \tag{3.50}$$

The following theorem shows that the sum f of a convergent power series is not only analytic at its center z_0, but at any point z_1 in its disk of convergence. Moreover it tells us how the coefficients of the power series at z_1 can be determined from the coefficients at z_0.

Theorem 3.3.1 (Weierstrass Rearrangement Theorem)**.** *The sum of a power series is analytic at any point in its disk of convergence. If f is given by (3.50) for $|z-z_0| < r$, and z_1 satisfies $|z_1 - z_0| < r$, then*

$$f(z) = b_0 + b_1\,(z-z_1) + \ldots + b_k\,(z-z_1)^k + \ldots, \qquad |z-z_1| < r_1, \tag{3.51}$$

where $r_1 := r - |z_1 - z_0|$ and the coefficients b_k are given by the convergent series

$$b_k = \sum_{n=k}^{\infty} \binom{n}{k} a_n (z_1 - z_0)^{n-k}, \qquad k = 0, 1, \ldots. \tag{3.52}$$

Proof. Let $|z - z_1| < r_1$. Substituting $z - z_0 = (z - z_1) + (z_1 - z_0)$ into (3.50), we obtain

$$f(z) = \sum_{n=0}^{\infty} a_n \left((z-z_1) + (z_1-z_0)\right)^n = \sum_{n=0}^{\infty} a_n \sum_{k=0}^{n} \binom{n}{k} (z-z_1)^k\,(z_1-z_0)^{n-k}.$$

In order to prove the assertion, it only remains to change the order of summation in the double series. By Theorem 2.6.4, it suffices to show that this series converges *absolutely*. To this end we remark that

$$\sum_{n=0}^{\infty} |a_n| \sum_{k=0}^{n} \binom{n}{k} |z-z_1|^k\,|z_1-z_0|^{n-k} = \sum_{n=0}^{\infty} |a_n| \left(|z-z_1| + |z_1-z_0|\right)^n.$$

The last sum converges because $|z - z_1| + |z_1 - z_0| < r$, so that the power series (3.50) converges absolutely at the point $z = z_0 + |z - z_1| + |z_1 - z_0|$. ☐

The radius r_1 has the largest value for which the disk $\{z : |z-z_1| < r_1\}$ is contained in the disk $\{z : |z-z_0| < r\}$. Note that r may be smaller than the radius of convergence R of the power series (3.50). More important is the observation that the power series (3.51) can have a radius of convergence which is even *larger* than $R - |z_1|$. If this happens, the rearranged series (3.51) converges at points outside the disk $|z-z_0| < R$, where the original function f was not even defined! This will allow one to *extend* f to a larger domain. We shall investigate this method of *analytic continuation* in detail in Section 3.6.

Example 3.3.1 (Binomial Series). The binomial series with parameter $\alpha = 1/2$,

$$f(z) = \sum_{k=0}^{\infty} \binom{1/2}{k} (z-1)^k, \tag{3.53}$$

has radius of convergence 1 (see Example 3.2.1) and its disk of convergence contains the point $z_1 = \mathrm{e}^{\mathrm{i}\pi/4}$. The rearrangement of (3.23) with center z_1 is

$$\mathrm{e}^{\mathrm{i}\pi/8} \sum_{k=0}^{\infty} \mathrm{e}^{-\mathrm{i}k\pi/4} \binom{1/2}{k} (z-z_1)^k \tag{3.54}$$

and has radius of convergence 1. While (3.53) diverges at $z = \mathrm{i}$, the rearranged series (3.54) converges at that point.

Definition 3.3.2. A complex function $f : D \subset \mathbb{C} \to \mathbb{C}$ is said to be *analytic on* A (or analytic in A) if A is a subset of D and f is analytic at every point of A. We say that f is *analytic* if it is *analytic* on its domain set. A function which is analytic on the entire complex plane is called *entire*.

Lemma 3.3.3. *For any complex function $f : D \subset \mathbb{C} \to \mathbb{C}$ the set A_f of all points in D at which f is analytic is open.*

Proof. If A_f is empty, there is nothing to prove. If $z_0 \in A_f$, then f has a Taylor expansion at z_0 which converges in a open disk D_0 centered at z_0. By Theorem 3.3.1, $D_0 \subset A_f$. □

It follows from Theorem 3.2.3 that every analytic function is continuous. Theorem 3.2.7 tells us that sum, difference, and product of two analytic functions f and g are analytic on D, while the quotient f/g is analytic on $\{z \in D : g(z) \neq 0\}$. Also the composition $g \circ f$ of analytic functions $f : D_f \to D_g$ and $g : D_g \to \mathbb{C}$ is analytic on D_f.

Polynomials, the exponential function, and the trigonometric functions sine and cosine are entire. A rational function f is analytic on $\{z \in \mathbb{C} : f(z) \neq \infty\}$.

In Section 3.5 we shall extend Definition 3.3.2 to admit functions which are defined on a subset of the Riemann sphere and may attain infinite values.

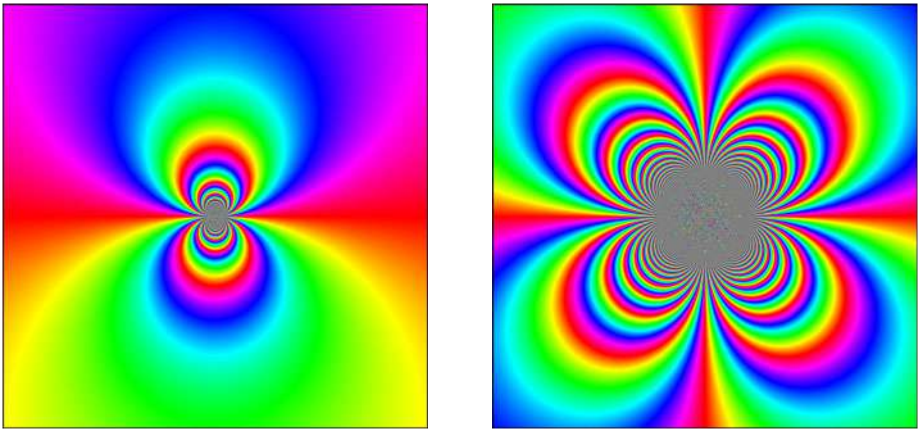


Figure 3.21: Phase portraits of $\exp(1/z)$ and $\exp(1/z^2)$

Example 3.3.2. The function f defined on the punctured plane $\dot{\mathbb{C}} := \mathbb{C} \setminus \{0\}$ by $f(z) = \exp(1/z)$ is analytic by Theorem 3.2.11. The function cannot be extended to an analytic function at $z_0 = 0$, since the extended function would be continuous there, but $f(z) \to \infty$ if $z \in \mathbb{R}_+$ and $z \to 0$, while $f(z) \to 0$ if $-z \in \mathbb{R}_+$ and $z \to 0$.

Figure 3.21 shows the phase portraits of $\exp(1/z)$ and $\exp(1/z^2)$ in the square defined by $|\operatorname{Re} z| < 1$, $|\operatorname{Im} z| < 1$. The wild behavior near the origin is typical for functions with a so-called *essential singularity*, which will be studied in Section 4.4. The curious reader is invited to look for possible reasons right away.

Example 3.3.3 (Jacobi Theta Function). An interesting family of entire functions are the *Jacobi Theta functions*,[4] given by the series

$$\vartheta(z) := \sum_{k=-\infty}^{\infty} q^{k^2} \mathrm{e}^{2k\pi \mathrm{i}\, z}, \qquad z \in \mathbb{C},$$

where q is a complex parameter with modulus less than one. In order to show that ϑ is entire, we consider the power series

$$f(z) := \sum_{k=1}^{\infty} q^{k^2} z^k = q\,z + q^4 z^2 + q^9 z^3 + q^{16} z^4 + \dots .$$

This series converges for all $z \in \mathbb{C}$, because

$$\limsup_{k\to\infty} \sqrt[k]{|q|^{k^2}} = \limsup_{k\to\infty} |q|^k = 0,$$

---

[4] Note that there are several closely related functions called Jacobi Theta functions.

and thus the function f is entire. The function g defined by $g(z) := \exp(2\pi\mathrm{i}\, z)$ is also entire and has no zeros in $\mathbb{C}$, so that its reciprocal $1/g$ is also entire. Finally,

$$\vartheta(z) = 1 + f\big(g(z)\big) + f\big(1/g(z)\big), \qquad z \in \mathbb{C}.$$

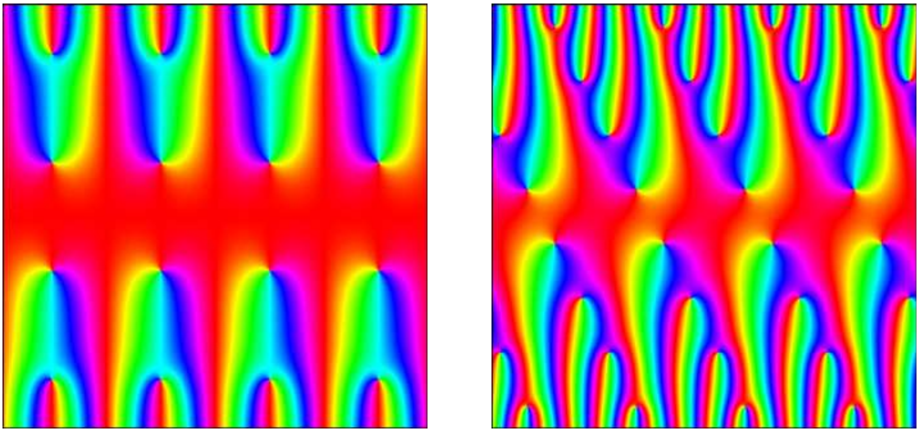


Figure 3.22: Phase portraits of two Jacobi Theta functions

The function g, and consequently ϑ, is periodic with period 1. The parameter q is said to be the *nome* of the Theta function. It is often represented as $q = \exp(\mathrm{i}\pi\tau)$, where τ is a complex number with $\operatorname{Im}\tau > 0$.

Figure 3.22 shows the phase portraits of two Jacobi Theta functions with parameters $\tau = \mathrm{i}$ and $\tau = -1/4 + \mathrm{i}/2$ in the square $|\operatorname{Re} z| < 2$, $|\operatorname{Im} z| < 2$.

Local Normal Forms. Our next goal is to generalize the factorization theorem for polynomials.

Theorem 3.3.4 (Local Normal Form)**.** *Let $f : D \subset \mathbb{C} \to \mathbb{C}$ be analytic on D. If f is not constant in a neighborhood of $z_0 \in D$, then there exist a positive integer m and an analytic function $g : D \subset \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ with $g(z_0) \neq 0$ such that for all $z \in D$*

$$f(z) = f(z_0) + (z - z_0)^m\, g(z). \tag{3.55}$$

The integer m and the function g are uniquely determined.

Proof. Assume that the Taylor series $f(z) = \sum a_k\,(z - z_0)^k$ of f at z_0 converges in a disk D_0. Denoting by a_m the first non-zero coefficient among $a_1, a_2, a_3 \ldots$, we have

$$f(z) = f(z_0) + (z - z_0)^m \sum_{k=m}^{\infty} a_k\,(z - z_0)^{k-m}, \quad z \in D_0.$$

The sum $g_0(z)$ of the series $\sum_{k=m}^{\infty} a_k\,(z-z_0)^{k-m}$ is an analytic function in D_0 with $g_0(z_0)=a_m\neq 0$. The function g defined in D by

$$g(z) := \begin{cases} (z-z_0)^{-m}\,\big(f(z)-f(z_0)\big) & \text{if} \quad z\in D\setminus\{z_0\} \\ a_m & \text{if} \quad z=z_0 \end{cases}$$

is analytic in $D\setminus\{z_0\}$. Since it coincides with g_0 in D_0 it is also analytic at z_0.

For proving uniqueness we assume that $(z-z_0)^n g_1(z)=(z-z_0)^m g_2(z)$ with $n>m$ for all $z\in D$. Then $(z-z_0)^{n-m}g_1(z)=g_2(z)$, and the left-hand side vanishes at z_0 while $g_2(z_0)\neq 0$. So $m=n$ and then $g_1=g_2$ is obvious. ☐

Order of a Function at a Point. The normal form in Theorem 3.3.4 describes the local behavior of an analytic function near some point z_0. The crucial parameter in this representation is the integer m.

Definition 3.3.5. The integer m in the representation (3.55) is called the *order* (or the *multiplicity*) of the function f at z_0 and denoted by $\operatorname{ord}(f,z_0)$. If f is constant in a neighborhood of z_0 we set $\operatorname{ord}(f,z_0):=\infty$. If, in particular, $f(z_0)=0$, then m is said to be the *order* (or *multiplicity*) of the zero z_0.

The order $m=\operatorname{ord}(f,z_0)$ of a function at z_0 can be easily read off from its phase portrait. Figure 3.23 shows three typical portraits with enhanced isochromatic lines (contour lines of the phase). The pictures on the left and in the middle correspond to cases where $f(z_0)\neq 0$. The case $m=1$ is shown on the left. If m is greater than one, we see a *color saddle*; the picture for $m=3$ is shown in the middle. In both cases the number of isochromatic passing through z_0 is equal to m, that is, $2m$ isochromatic rays emerge from z_0.

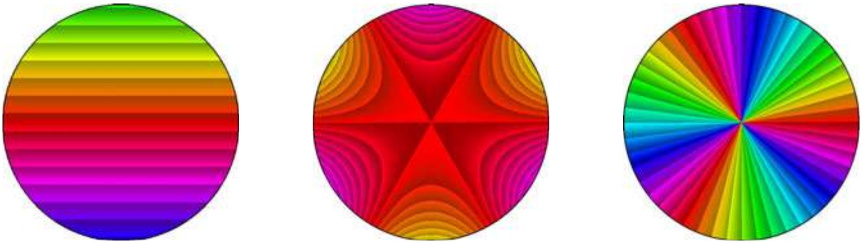


Figure 3.23: Enhanced phase portraits near z_0 corresponding to the normal forms

If z_0 is a zero of order m, we get the phase portrait on the right-hand side. Here isochromatic lines of all colors meet at z_0 and the number of such lines of one specific color is equal to m. For a simple explanation of these observations we remark that $g(z)$ is almost constant and close to $g(z_0)$ if z is close to z_0. It is easily seen that then the phase of f does not change, and the phase portraits are only slightly disturbed, if g is replaced by $g(z_0)$. We do not work this out here; a detailed description of the (local and global) structure of isochromatic lines will be given in Volume 2 (see also [69]).

A Property of Zeros and a-Values. As an immediate corollary of Theorem 3.3.4 we get the following result which shows, in particular, that all zeros of non-constant analytic functions are isolated.

Lemma 3.3.6. *If f is analytic at z_0 and $a := f(z_0)$, then there exists a disk D_0 with center z_0 such that either $f(z) = a$ for all $z \in D_0$ or $f(z) \neq a$ for all $z \in D_0 \setminus \{z_0\}$.*

Note that this statement does *not* claim that the zero set of an analytic function has no accumulation points. It just tells us that an accumulation point of zeros cannot be a zero itself. Since analytic functions are continuous, this in turn implies that such accumulation points cannot lie in the domain of f (but of course they must belong to the closure of the domain).

3.4 Analytic Functions in Planar Domains

Equipped with the necessary prerequisites, we will now derive several basic results about analytic functions. And we shall discover that some of these properties are really astonishing!

As we have already seen in Section 3.3, it is natural to require that the domain set D of an analytic function is open. From now on, we shall also assume that D is connected, i.e., D is a *domain* in the sense of Definition 2.7.1. This assumption is not too strong, since any open set in $\mathbb{C}$ is the disjoint union of domains, but it simplifies life a lot. In particular it is important when local statements about power series will be "lifted" to global results for analytic functions. This will be demonstrated in the proof of the following theorem.

Theorem 3.4.1 (Identity Theorem, Uniqueness Principle)**.** *Let f and g be analytic functions in a domain D. If there exists a sequence $(z_n) \subset D \setminus \{z_0\}$ such that $z_n \to z_0 \in D$ and $f(z_n) = g(z_n)$ for all $n \in \mathbb{Z}_+$, then $f(z) = g(z)$ for all $z \in D$.*

Proof. 1. The function $h := f - g$ has a sequence of zeros which converge to $z_0 \in D$. Continuity of h implies that $h(z_0) = 0$, so that z_0 is a zero of h which is not isolated. Since h is analytic in D, we infer from Lemma 3.3.6 that $h(z) = 0$ in some disk D_0 with center z_0.

2. We pick any point z_1 in D and show that $h(z_1) = 0$. By Proposition 2.7.13, there is a path $\gamma : [0,1] \to D$ from z_0 to z_1. Then the set

$$S := \big\{ s \in [0,1] : h\big(\gamma(t)\big) = 0 \text{ for all } t \in [0,s] \big\}$$

is not empty and we denote by s_0 its supremum. Continuity of h implies that $h\big(\gamma(s_0)\big) = 0$. Since $h\big(\gamma(t)\big) = 0$ for all $t \in [0, s_0]$, Lemma 3.3.6 tells us that $h(z) = 0$ in a neighborhood of $\gamma(s_0)$. This is only possible if $s_0 = 1$, because otherwise $h\big(\gamma(t)\big) = 0$ for all t in an interval $[0, s_1]$ with $s_1 > s_0$. □

Zeros of Analytic Function. The last theorem establishes the surprising fact that a function which is analytic in a domain is completely determined by its values in an arbitrarily small disk or even on a Jordan arc. We state another result concerning the zeros of such a function.

Corollary 3.4.2. *If $f \neq 0$ is analytic in a domain D and K is a compact subset of D, then the number of zeros of f in K is finite.*

Proof. If f had infinitely many zeros in K, there would exist a sequence (z_k) of such zeros which converge to a point $z_0 \in K \subset D$. But then $f = 0$ on D by Theorem 3.4.1. ☐

Nevertheless an analytic function $f \neq 0$ can have infinitely many zeros in D. If this happens, the zeros must have an accumulation point z_0 on $\widehat{\mathbb{C}}$. Since z_0 cannot lie in D, it must be on the boundary of D (considered as a subset of $\widehat{\mathbb{C}}$). For an entire function the only possible accumulation point of zeros is the point at infinity. An example of this type is the sine function. Substituting the variable z by $1/z$ we obtain the following example.

Example 3.4.1. The function $\sin(1/z)$ is analytic in $\mathbb{C} \setminus \{0\}$ and has the zeros $z_k = 1/(k\pi)$ with $k = \pm 1, \pm 2, \ldots$, which accumulate at the origin. Figure 3.24 (left) depicts this function in the square $|\operatorname{Re} z| < 1/2$, $|\operatorname{Im} z| < 1/2$.

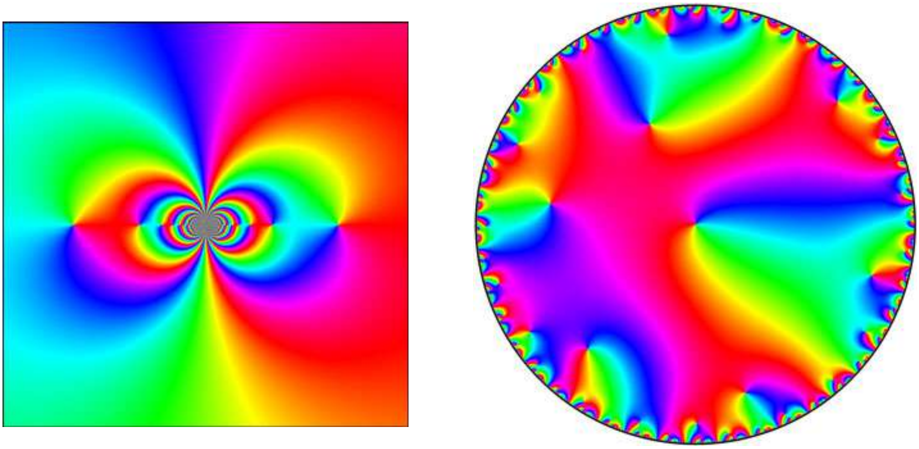


Figure 3.24: Two analytic functions with an infinite number of zeros

The phase portrait on the right gives an impression[5] of a function for which every point on the boundary of the domain is an accumulation point of zeros. It belongs to the class of *infinite Blaschke products*, which we shall study in Section 5.4. The function depicted has zeros $z_k = \left(1 - k^{-1.01}\right) \mathrm{e}^{\mathrm{i}\,k}$ with $k = 1, 2, \ldots$.

---

[5] What we see is of course an approximation, a printed phase portrait has a finite number of pixels and is unable to depict infinitely many zeros.

Extremal Values. The theorem on normal forms opens the door to a number of structural theorems about analytic functions. The next result generalizes the maximum and minimum principle for polynomials.

Theorem 3.4.3 (Maximum and Minimum Principle)**.** *Let* $f : D \subset \mathbb{C} \to \mathbb{C}$ *be a non-constant analytic function. Then* $|f|$ *has no local maximum in* D*, and every local minimum of* $|f|$ *is a zero of* f*.*

Proof. Assume that $|f|$ attains a (finite) maximum or minimum at $z_0 \in D$. By Theorem 3.4.1 f is not locally constant, so that we can apply Theorem 3.5.5 and write

$$f(z) = f(z_0) + (z - z_0)^m \, g(z),$$

where g is analytic in D and $g(z_0) \neq 0$. The rest of the proof is similar to that of Theorem 3.1.4 on page 65. ☐

The maximum principle can also be rephrased as follows: if f is analytic in a domain D and $|f|$ attains a local maximum at $z_0 \in D$, then f is constant in D.

The Argument Principle. We turn next to a central result (Theorem 3.4.5) which, besides being beautiful and useful, is also a prime example of the interplay of geometry and analysis. But first some more preparation.

Assume that f is a continuous complex function (not necessarily analytic) on D. If γ is a path in D, then the composition $f \circ \gamma$ is a path in the image set $f(D)$. If γ is a *closed* path and f has no zeros on its trace $[\gamma]$, then the image path $f \circ \gamma$ is also closed and its winding number $\operatorname{wind}_\gamma f := \operatorname{wind}(f \circ \gamma)$ is well defined.

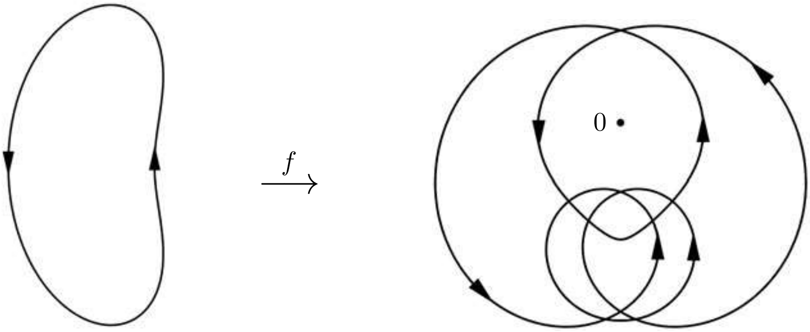


Figure 3.25: A simple closed path γ and its image $f \circ \gamma$ with $\operatorname{wind}(f \circ \gamma) = 2$

Figure 3.25 shows the traces of a simple closed path γ and its image $f \circ \gamma$. When the point z travels along $[\gamma]$, then $f(z)$ traverses $f([\gamma])$. The winding number $\operatorname{wind}_\gamma f$ of the image path $f \circ \gamma$ equals two.

Lemma 3.4.4 (Winding Numbers)**.** *Let* γ *be a closed path in the domain* D *and assume that* $f, g : D \to \mathbb{C}$ *are continuous complex functions which have no zeros* *on the trace of* γ*. Then*

$$\operatorname{wind}_\gamma (f\,g) = \operatorname{wind}_\gamma f + \operatorname{wind}_\gamma g.$$

Proof. If a_f and a_g are continuous branches of the argument along $f \circ \gamma$ and $g \circ \gamma$ (see Lemma 2.7.19), then $a_f + a_g$ is a continuous branch of the argument along $(f\,g) \circ \gamma$. Now the result follows from the definition of the winding number (Definition 2.7.20). ☐

Example 3.4.2. If $\gamma \subset \mathbb{C} \setminus \{z_0\}$ with $\operatorname{wind}(\gamma, z_0) = 1$ and $f(z) := (z - z_0)^n$ with $n \in \mathbb{Z}$, then $\operatorname{wind}_\gamma f = n$.

For *analytic* functions, there is an amazing relationship between the number of zeros of f inside a Jordan curve $J = [\gamma]$ and the winding number of the image path $f \circ \gamma$. Miraculously, the values of f on the curve J determine how many zeros f has in the interior of J. In order to formulate the result we define

$$\operatorname{wind}_J f := \operatorname{wind}(f \circ \gamma),$$

where γ is a positively oriented parametrization of J. Note that this definition is independent of the choice of γ.

Theorem 3.4.5 (Argument Principle for Analytic Functions)**.** *Let* $f : D \to \mathbb{C}$ *be analytic in the domain* D*, and let* J *be a positively oriented Jordan curve with* $J \cup \operatorname{int} J \subset D$*. If* f *has no zeros on* J*, then the number of zeros of* f *in the interior of* J*, counted with multiplicity, is equal to* $\operatorname{wind}_J f$*.*

Proof. Let G denote the interior of J. Then $K := G \cup J \subset D$ is compact, and from Corollary 3.4.2 it follows that the number of zeros of f in K is finite. By assumption all these zeros $z_1, \dots, z_k$ belong to G. Denoting by $m_1, \dots, m_k$ their multiplicities, repeated application of Theorem 3.5.5 yields the representation

$$f(z) = (z - z_1)^{m_1} \cdot \ldots \cdot (z - z_k)^{m_k}\, g(z), \tag{3.56}$$

where g is analytic in D and has no zeros in K. Let $f_j : z \mapsto (z - z_j)^{m_j}$ for $j = 1, \dots, k$. The functions g and f_j do not vanish on J, and hence the winding numbers of $g \circ \gamma$ and $f_j \circ \gamma$ are well defined for $j = 1, \dots, k$. By Lemma 3.4.4 we have

$$\operatorname{wind}_\gamma f = \operatorname{wind}_\gamma g + \operatorname{wind}_\gamma f_1 + \ldots + \operatorname{wind}_\gamma f_k.$$

From Example 3.4.2 we derive that $\operatorname{wind}_\gamma f_j = m_j$, so it only remains to show that $\operatorname{wind}_\gamma g = 0$.

Let z_0 be any point in G. By Lemma 2.7.26, γ is freely homotopic in K to z_0, i.e., there is a continuous family of paths γ_s $(0 \le s \le 1)$ such that $\gamma_0 := \gamma$, $\gamma_1 = \text{const} = z_0$ and $\gamma_s \subset K$ for all s.

The homotopy $s \mapsto \gamma_s$ induces a homotopy $s \mapsto g \circ \gamma_s$ of $g \circ \gamma$ to the constant path $g(z_0)$. Since $g \circ \gamma_s \subset \mathbb{C} \setminus \{0\}$, we infer from Lemma 2.7.22 that $\operatorname{wind}_\gamma g_s = 0$ for all s, which completes the proof. ☐

Remark 3.4.6. Assume that f is analytic in D, that J is a positively oriented Jordan curve in D, and that $a \in \mathbb{C} \setminus f(J)$. Then $\mathrm{wind}_J\,(f-a)$ counts the number of solutions of the equation $f(z) = a$ in the interior of J. In particular, this winding number is *always non-negative.*

Seeing the Argument Principle. The argument principle has a nice interpretation in the phase portrait. To describe it, we first translate the definition of winding number into the language of colors. Let $\gamma : [0,1] \to D$ be a closed path, and assume that $f : D \to \mathbb{C}$ has no zeros on $[\gamma]$. We let z traverse γ and observe the point $f(z)$ moving along the image path $f \circ \gamma$ in the phase portrait of f. Any such point $f(z)$ caries a color $c(z) := f(z)/|f(z)| = \exp\big(\mathrm{i}\arg f(z)\big)$ of the color circle. If $\arg f(z)$ is chosen as a continuous branch along γ, then $c\big(\gamma(1)\big) - c\big(\gamma(0)\big) = 2\pi\,\mathrm{wind}_\gamma f$. So, in order to determine $\mathrm{wind}_\gamma f$ we just count how many times the color of the point $\gamma(t)$ in the phase portrait of f rotates around the color circle when $\gamma(t)$ runs through $[\gamma]$ once in positive direction.

This is illustrated in Figure 3.26, where we see a simple closed path γ in the phase portrait of f on the left-hand side, and its image path $f \circ \gamma$ in the complex w-plane, which is colored according to the phase of w.

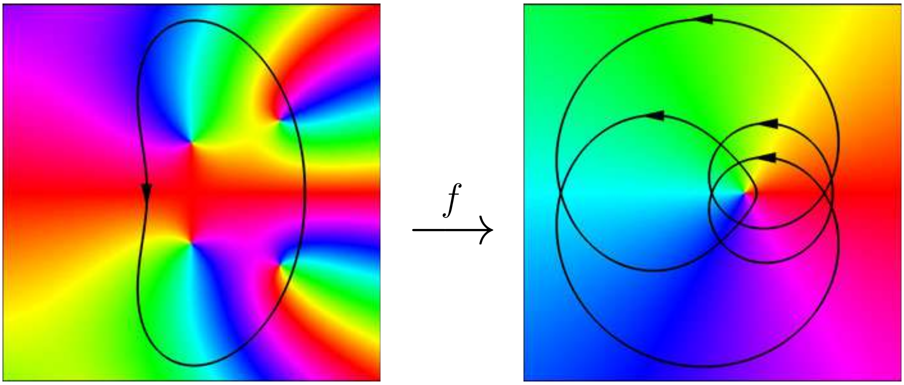


Figure 3.26: A path γ with chromatic number $\mathrm{chrom}_f\,\gamma = 4$ and its image

Chromatic Number of a Path. Sometimes it is useful to reverse the roles of f and γ. Then we call $\mathrm{wind}_\gamma f$ the *chromatic number* of γ in the phase portrait of f and denote it by $\mathrm{chrom}_f\,\gamma$. If J is a Jordan curve, we choose a positively oriented parametrization γ and set $\mathrm{chrom}_f\,J := \mathrm{chrom}_f\,\gamma$. This definition is independent of the choice of γ. The chromatic number of the Jordan curve J shown on the left in Figure 3.26 is equal to four.

The above interpretation of the argument principle is not only intuitive, but may bring forth a couple of new ideas. Looking at Figures 3.26 or 3.27 in search of zeros of f, one almost automatically follows the isochromatic lines (for instance the yellow ones), expecting that every such line ends up at a zero. A detailed investigation of this idea is deferred to Volume 2 (see also [69]), but it might be helpful to have an informal discussion right now.

In fact, starting at a yellow point on J and moving along the corresponding yellow line (with red to the right and green to the left), we may cross the curve J several times, but typically we are eventually trapped either in $\operatorname{int} J$ or $\operatorname{ext} J$. Here "typically" means that there may be exceptions, but these disappear if an isochromatic line of slightly different color is chosen. Any isochromatic line which finally stays in $\operatorname{int} J$ typically ends up at a zero of f, so that the isochromatic lines indeed guide us to the zeros.

This observation and the argument principle become even more transparent if we reverse the direction of our journey along an isochromatic line. Instead of travelling from a point on J towards a zero, we start at a zero in $\operatorname{int} J$. Then, with the possible exception of just a *finite number* of isochromatic lines, we *always* arrive at a point on J.

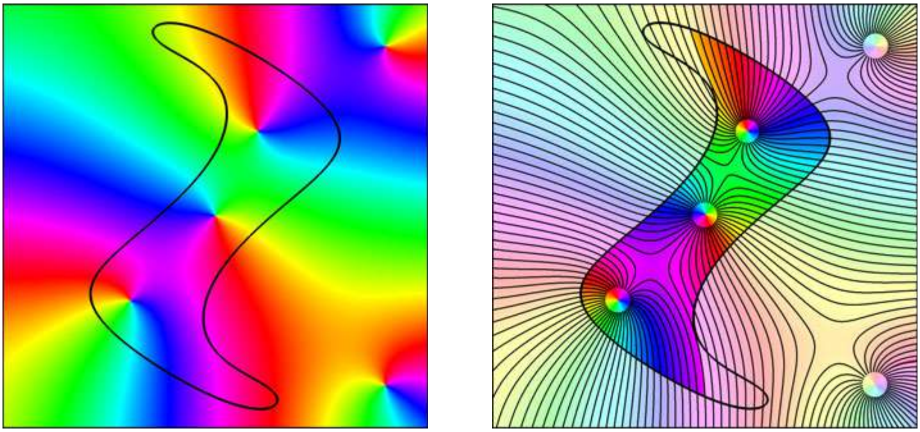


Figure 3.27: Phase flow along the isochromatic lines of an analytic function

The Phase Flow. Considering the movement along the isochromatic lines *simultaneously* for all points, gives the impression of a *flow*. The streamlines of that flow are the isochromatic lines, and one can imagine that it carries a colored substance, which we call "phase". The "amount of phase" which is created by a zero z_0 of order m is $2\pi m$.

The "phase flow" transports "phase" from the zeros of f through the domain $G := \operatorname{int} J$ to its boundary J where it is deposited as soon as it reaches the curve J. Note that some parts of J may remain empty, because they lie in the "shadow" of other parts (see Figure 3.27). Now the amount of phase deposited at J must be equal to the amount of phase emerging in G, which is equal to 2π times the number n of zeros of f in G (counted with multiplicity).

On the other hand, the amount of phase deposited at J is equal to the *chromatic number* $\operatorname{chrom}_f J$. To make this plausible, we choose a positively oriented parametrization γ of J and consider an arc $J_0 \subset J$ with initial point $a = \gamma(t_1)$ and terminal point $b = \gamma(t_2)$ $(0 \le t_1 < t_2 \le 1)$. The amount of phase which flows through J_0 (assuming, for the moment, that it is not deposited on J) from the interior of J to the exterior of J is equal to $\arg\big(f(b)\big) - \arg\big(f(a)\big)$, where arg denotes a *continuous branch* of the argument along γ (see Lemma 2.7.19). In particular we have

$$2\pi\, n = \arg\,\big(f(\gamma(1))\big) - \arg\,\big(f(\gamma(0))\big) = 2\pi \operatorname{chrom}_f \gamma = 2\pi \operatorname{chrom}_f J.$$

We point out that the above interpretation leads to a *quantitative version* of the argument principle, which tells us how "phase" emerging from the zeros of the function is distributed along the Jordan curve J. We shall explore these ideas further when we study *"phase diagrams"* in Volume 2.

Rouché's Theorem. The next application of the argument principle shows that the number of zeros of f inside a Jordan curve J is stable with respect to small perturbations of f on the curve J and gives an explicit estimate of by how much f can be disturbed. The following elegant version of this result is symmetric with respect to the involved functions and slightly more general than the usual one, which assumes the stronger condition $|f(z) - g(z)| < |f(z)|$ instead of inequality (3.57).

Theorem 3.4.7 (Rouché)**.** *Let D be a domain and assume that the positively oriented Jordan curve J and its interior are contained in D. If f and g are analytic in D and*

$$|f(z) - g(z)| < |f(z)| + |g(z)| \tag{3.57}$$

for all $z \in J$, then f and g have the same number of zeros in $\operatorname{int} J$*, counting multiplicities.*

Proof. It follows from (3.57) that $f(z) \ne 0$ and $g(z) \ne 0$ for all $z \in J$. For $s \in [0,1]$, we define the analytic functions h_s by $h_s(z) := f(z) + s\big(g(z) - f(z)\big)$ and show that $h_s(z) \ne 0$ for all $z \in J$ and $0 < s < 1$. Indeed, if $h_s(z) = 0$, then $s\,g(z) = (s-1)f(z)$, so that $g(z)/f(z) = (s-1)/s$ is negative. Consequently, contrary to (3.57), we have

$$\left|1 - \frac{g(z)}{f(z)}\right| = 1 + \left|\frac{g(z)}{f(z)}\right|.$$

The mapping $h : [0,1]\times[0,1] \to \mathbb{C}\setminus\{0\}$ defined by $h(s,t) := h_s(\gamma(t))$ is continuous. Since $h_0 = f$ and $h_1 = g$, this shows that $f \circ \gamma$ and $g \circ \gamma$ are freely homotopic in $\mathbb{C}\setminus\{0\}$ and hence $\operatorname{wind}_\gamma f = \operatorname{wind}_\gamma g$. Now the assertion follows from the argument principle. □

The Open Mapping Principle. Another interesting property of analytic functions which we study next is basically a statement about the *local* behavior of non-constant analytic functions. Though it could have been stated already in Section 3.3, we present a global version here.

It is a well-known fact (and easy to prove) that the *preimage* $f^{-1}(V)$ of an open set V is always open, provided that f is a *continuous* function. Though the corresponding statement is not true, in general, for the *image* $f(U)$ of an open set U, it is remarkable that all non-constant *analytic* functions have this property.

Theorem 3.4.8 (Open Mapping Principle)**.** *Let $f : D \to \mathbb{C}$ be analytic and non-constant in the domain D. If $U \subset D$ is an open set, then $f(U)$ is an open set.*

Proof. Intuitively the result follows directly from the argument principle: if a closed path has a positive winding number about some point w_0, then this also holds for all points w in a sufficiently small neighborhood of w_0.

We convert this into a strict proof using Rouché's theorem: if $w_0 \in f(U)$, then there exists $z_0 \in D$ with $w_0 = f(z_0)$. Since U is open, some closed disk K centered at z_0 is contained in U. By Corollary 3.4.2, z_0 is an *isolated* zero of $f - w_0$, i.e., the radius of K can be chosen so small that z_0 is the only point in K with $f(z) = w_0$. Then the absolute value of $f - w_0$ has a positive minimum δ on the boundary J of K. If $w \in \mathbb{C}$ satisfies $|w - w_0| < \delta$, then, by Rouché's theorem, the functions $f - w_0$ and $f - w$ have the same number of zeros in K. Since $f(z_0) = w_0$, this number is positive, and thus all points w with $|w - w_0| < \delta$ are contained in the image $f(U)$. □

The open mapping principle sheds new light on the maximum/minimum principle: for any $z_0 \in D$ the image $f(U)$ of any neighborhood U of z_0 contains a neighborhood of $f(z_0)$, so that $|f|$ can neither have a local maximum nor a positive local minimum at z_0.

A Uniqueness Result for Phase Portraits. In order to ascertain to which extent an analytic function is determined by its phase portrait, we state another immediate consequence of the open mapping principle.

Corollary 3.4.9. *If f is analytic in a domain D and if one of the functions $|f|$, $\operatorname{Re} f$, or $\operatorname{Im} f$ is constant on an open subset of D, then f is constant on D.*

The following result tells us that analytic functions with the same phase portrait in some domain are essentially the same. So phase portraits indeed provide *analytic* functions with an *individual* face.

Theorem 3.4.10. *If f and g are analytic in the domain D and have the same phase (portrait) on an open set $U \subset D$, then g is a positive multiple of f, i.e., $g(z) = c\, f(z)$ for all $z \in D$ and some positive constant c.*

Proof. The assertion is obvious if f is the zero function. Otherwise we can find a disk $V \subset U$ which contains no zeros of f. Then the quotient g/f is analytic in V, and writing $f = |f|\,\psi(f)$ and $g = |g|\,\psi(g)$ with $\psi(f) = \psi(g)$, we see that g/f

attains only positive real values. Then g/f must be a positive constant on D, by virtue of Corollary 3.4.9. $\square$

Solving Equations. Rouchés theorem yields an interesting by-product concerning the solvability of equations $f(z) = w$.

Theorem 3.4.11. *Let f be analytic in the domain D, $z_0 \in D$, $w_0 := f(z_0)$, and assume that* $\operatorname{ord}(f, z_0) = m \neq \infty$, *i.e., z_0 is a zero of $f - w_0$ with multiplicity m. Then the following statements are true.*

(i) *There exist positive numbers ε and δ such that for all w with $0 < |w - w_0| < \delta$ the equation $f(z) = w$ has exactly m isolated solutions with $0 < |z - z_0| < \varepsilon$.*

(ii) *If $m = 1$ then for all sufficiently small positive r the restriction of f to the disk $U := D_r(z_0)$ is a homeomorphism of U onto a simply connected domain $V := f(U)$.*

Proof. 1. By Lemma 3.3.6, the point z_0 is an isolated zero of $f - w_0$, so that we can choose a disk $D_0 \subset D$ centered at z_0 which contains no other zero of $f - w_0$. Since $|f - w_0|$ has a positive minimum c on the boundary of D_0, it follows from Theorem 3.4.7 that the function $f - w$ has exactly m zeros (counted with multiplicity) in D_0 for all $w \in \mathbb{C}$ with $|w - w_0| < c$.

2. The Taylor series $f(z) = b_0 + b_1 (z - z_1) + b_2 (z - z_1)^2 + \ldots$ of f at z_1 converges for all z_1 with $|z_1 - z_0| < \varepsilon$ if ε is sufficiently small. The coefficient b_1 depends on z_1 and is given by (3.52),

$$b_1(z_1) = \sum_{n=1}^{\infty} n\, a_n\, (z_1 - z_0)^{n-1}.$$

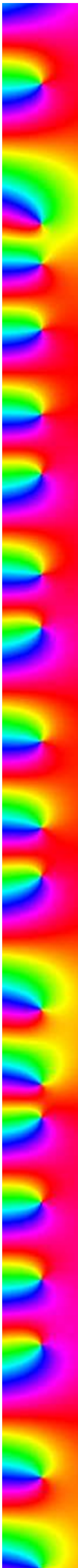
The sum on the right-hand side defines an analytic function of z_1 which is non-zero. Writing $b_1(z_1)$ in its normal form at z_0, we get

$$b_1(z_1) = a_1 + (z_1 - z_0)^k\, g(z_1),$$

where g is an analytic function of z_1 with $g(z_0) \neq 0$. Thus, whether $a_1 = 0$ or $a_1 \neq 0$, we conclude in both cases that $b_1(z_1) \neq 0$ if $0 < |z_1 - z_0| < \varepsilon$ and ε is sufficiently small. This implies $\operatorname{ord}(f, z_1) = 1$, so that all zeros of $f - f(z_1)$ are simple.

3. Let $m = 1$. We choose ε and δ according to (i). Since f is continuous at z_0, there exists a number r with $0 < r < \varepsilon$ such that the disk $U := D_r(z_0)$ is mapped into $D_\delta(w_0)$. By (i) f is injective on U, so it maps U bijectively onto $V := f(U)$. Let f^{-1} denote the inverse of $f : U \to V$. The open mapping principle guarantees that f^{-1} is continuous, so the mapping $f : U \to V$ is a *homeomorphism*, which implies that V is a simply connected domain. $\square$

The splitting of multiple zeros is illustrated in Figure 3.28. It shows the enhanced phase portraits of $f(z) = (\sin z)^5$, which has a zero of order 5 at $z_0 = 0$, and the function $f - 10^{-4}$ in the unit square $|\operatorname{Re} z| < 1$, $|\operatorname{Im} z| < 1$.



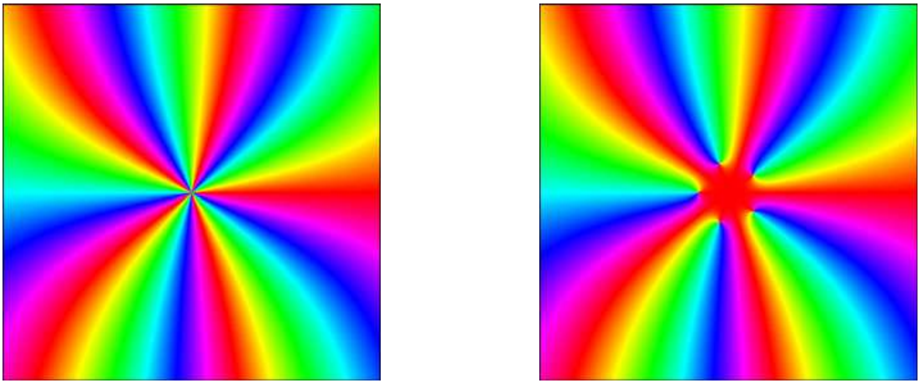


Figure 3.28: Splitting of a multiple zero under a small perturbation

The Schwarz Lemma. We close this section with a corollary of the maximum principle, which has so many applications and far reaching generalizations that whole books have been written about it (Avkhadiev and Wirths [3], see also Krantz [30], Chapter 2).

Theorem 3.4.12 (Schwarz Lemma). *Let f be analytic in the unit disk $\mathbb{D}$, and assume that $f(0)=0$ and $|f(z)|\le 1$ for all $z\in\mathbb{D}$. Then, for all $z\in\mathbb{D}$,*

$$|f(z)|\le |z|. \tag{3.58}$$

If $|f(z_0)|=|z_0|$ for some $z_0\in\mathbb{D}\setminus\{0\}$, then there exists a unimodular constant $c\in\mathbb{T}$ such that $f(z)=c\,z$ for all $z\in\mathbb{D}$.

Proof. The function g defined by $g(z):=f(z)/z$ is analytic in the punctured disk $D:=\mathbb{D}\setminus\{0\}$. In order to extend g to an analytic function in $\mathbb{D}$, we denote by $a_0+a_1\,z+a_2\,z^2+\dots$ the Taylor series of f at 0. Since $f(0)=0$, we have $a_0=0$. Setting $g(0):=a_1$, we obtain that $g(z)=a_1+a_2\,z+\dots$ for all z in a neighborhood of 0, i.e., g is also analytic at 0.

The estimate $|f(z)|\le 1$ guarantees that $|g(z)|\le 1/r$ if $0<|z|=r<1$. Referring to the maximum principle we conclude that $|g(z)|\le 1/r$ for $|z|\le r<1$. When r tend to 1, we get $|g(z)|\le 1$ for all $z\in\mathbb{D}$, which is equivalent to (3.58).

If $|f(z_0)|=|z_0|$ for some $z_0\in\mathbb{D}\setminus\{0\}$, then $|g(z_0)|=1$. Since $|g(z)|\le 1$ for all $z\in\mathbb{D}$, $|g|$ attains a maximum at $z_0\in\mathbb{D}$. By the maximum principle, g must be a constant (unimodular) function, and denoting this constant by c we find that $f(z)=cz$. ☐

The following example illustrates how the argument principle and Schwarz' lemma can be applied to analyze properties of analytic functions.

Example 3.4.3 (Blaschke Factors). The functions we are considering are special Möbius transformations, called *Blaschke factors*, which have the form

$$f(z) = c\,\frac{z - z_0}{1 - \overline{z}_0\, z}, \qquad |c| = 1,\ |z_0| < 1. \tag{3.59}$$

As is true for any Möbius transform, Blaschke factors map the extended complex plane $\widehat{\mathbb{C}}$ bijectively onto itself. The function f defined in (3.59) has a simple zero at z_0 and a simple pole at $1/\overline{z}_0$, and its inverse belongs to the same class,

$$f^{-1}(z) = \overline{c}\,\frac{z + z_0}{1 + \overline{z}_0 z}. \tag{3.60}$$

Evaluating the function f on the unit circle $\mathbb{T}$,

$$f\big(\mathrm{e}^{\mathrm{i}\varphi}\big) = c\,\frac{\mathrm{e}^{\mathrm{i}\varphi} - z_0}{1 - \overline{z}_0\mathrm{e}^{\mathrm{i}\varphi}} = c\,\mathrm{e}^{\mathrm{i}\varphi}\,\frac{1 - \mathrm{e}^{-\mathrm{i}\varphi} z_0}{\overline{1 - \mathrm{e}^{\mathrm{i}\varphi} z_0}},$$

we see that $|f(z)| = 1$ for all $z \in \mathbb{T}$. Since f is holomorphic in a neighborhood of $\mathbb{D}$ and not constant, it follows from the maximum principle that $|f(z)| < 1$ for all $z \in \mathbb{D}$, i.e., f maps the unit disk $\mathbb{D}$ into itself. Because f^{-1} has the same property, $f : \mathbb{D} \to \mathbb{D}$ is indeed a bijective mapping.

Conversely, any holomorphic function g which maps $\mathbb{D}$ bijectively onto itself is a Blaschke factor. To prove this claim, we set $z_0 := g^{-1}(0)$ and define f by (3.59) with $c = 0$. Then $h := f \circ g^{-1}$ and $h^{-1} = g \circ f^{-1}$ map $\mathbb{D}$ onto itself and satisfy $h(0) = h^{-1}(0) = 0$. By Schwarz lemma,

$$|h(z)| \le |z| = \big|h^{-1}\big(h(z)\big)\big| \le |h(z)|, \qquad z \in \mathbb{D},$$

so that $|h(z)| = |z|$ for all $z \in \mathbb{D}$. Invoking Schwarz' lemma again, we obtain that $h(z) = c\,z$ with $|c| = 1$. Finally,

$$\big(f \circ g^{-1}\big)(z) = h(z) = cz$$

for all $z \in \mathbb{D}$ implies that $f(z) = c\,g(z)$, so that g is indeed a Blaschke factor.

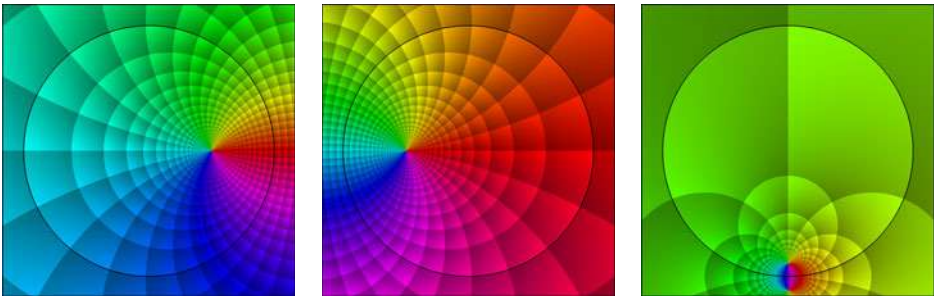


Figure 3.29: Phase portraits of three Blaschke factors

Figure 3.29 shows the phase portraits of three Blaschke factors with $c = 1$ and $z_0 = 0.5$, $z_0 = -0.5$, and $z_0 = -0.9\,\mathrm{i}$, respectively. The black line is the unit circle $\mathbb{T}$. In these pictures we see that the phase $\psi(f)$ changes monotonously along $\mathbb{T}$; this is a general property of Blaschke products. If γ denotes the standard parametrization of $\mathbb{T}$, then the argument principle gives

$$\operatorname{wind}(f \circ \gamma, w) = 1, \qquad w \in \mathbb{D}. \tag{3.61}$$

The chessboard coloring in Figure 3.30 is better suited to get an intuitive understanding of the manner in which a Blaschke factor maps the unit disk onto itself. Here the domain disk is depicted on the left, the range disk is seen on the right, and $z_0 = (1 + \mathrm{i})/2$.

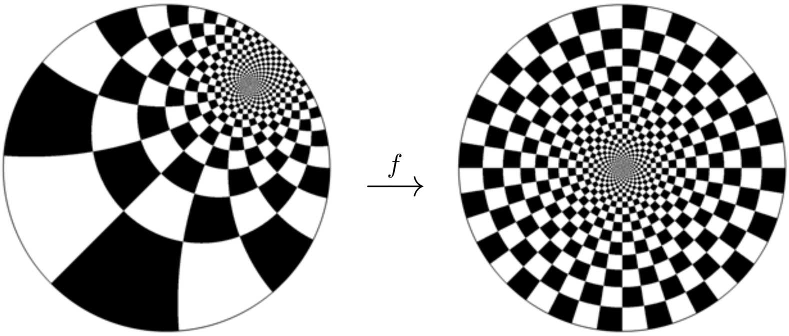


Figure 3.30: A Blaschke factor mapping the unit disk onto itself

Example 3.4.4 (Blaschke Products). As the name suggests, Blaschke factors are constituents of *Blaschke products*. A *finite* Blaschke product of *degree* (or order) n is a rational function of the form

$$f(z) = c \prod_{k=1}^{n} \frac{z - z_k}{1 - \overline{z}_k\, z}, \qquad |c| = 1, \quad |z_k| < 1 \qquad (k = 1, \ldots, n). \tag{3.62}$$

Since any factor $f_k(z) := (z - z_k)/(1 - \overline{z}_k\, z)$ is unimodular on the unit circle, the same holds for the product. The formula (3.61) for the winding number of the factors implies that

$$\operatorname{wind}(f \circ \gamma, w) = \operatorname{wind}(f_1 \circ \gamma, w) + \ldots + \operatorname{wind}(f_n \circ \gamma, w) = n, \quad w \in \mathbb{D}.$$

Consequently, by the argument principle, for every $w \in \mathbb{D}$ the equation $f(z) = w$ has exactly n solutions $z \in \mathbb{D}$ (counting multiplicity). This observation is a first hint that Blaschke products play a similar role in the unit disk as polynomials do in the complex plane. In fact these analogies go much deeper – in some sense, Blaschke products can be considered to be *"hyperbolic counterparts"* of polynomials.

Figure 3.31 visualizes a Blaschke product on the Riemann sphere, which has 50 randomly chosen zeros in the unit disk. Note that, due to stereographic projection, the colors appear in reversed order when we look at the sphere from the outside (see page 41 for an explanation). So the poles on the upper hemisphere pretend to be zeros, and the zeros on the lower hemisphere mimic poles.

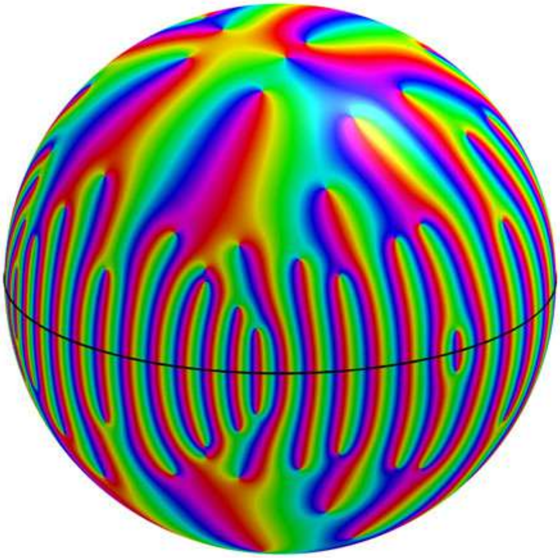


Figure 3.31: Phase portrait of a finite Blaschke product on the Riemann sphere

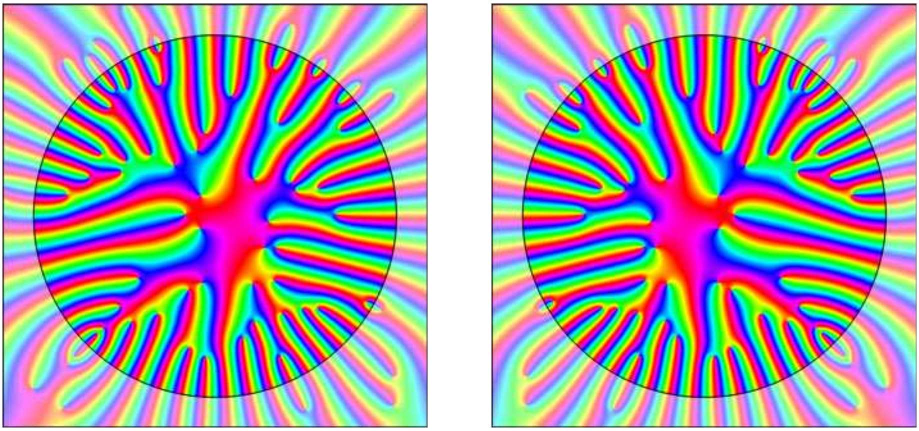


Figure 3.32: Phase portraits of a finite Blaschke product on the two hemispheres

The symmetry of the phase with respect to the equator which we observe in this picture can easily be verified. It is a consequence of a symmetry property of

Blaschke products with respect to reflection at the unit circle,

$$f\big(1/\overline{z}\big) = 1/\overline{f(z)}.$$

Because the phase of $1/\overline{w}$ is equal to the phase of w, this implies the symmetry property $\psi\big(f\big(1/\overline{z}\big)\big) = \psi\big(f(z)\big)$.

Figure 3.32 depicts the same function as in Figure 3.31 separately on the upper and the lower hemisphere. The orientation of the two squares has been chosen according to the convention in Section 2.5 (see page 41 for details), so that the two phase portraits are mirror symmetric with respect to the vertical axis separating them.

3.5 Analytic Functions on the Sphere

In this section we shall extend the concept of analyticity to functions defined on a subdomain D of the Riemann sphere $\widehat{\mathbb{C}}$. Moreover we shall allow that these functions attain infinite values.

General Definition of Analyticity. In order to include the point at infinity into the domain and the range of an analytic function we observe that the mapping $z \mapsto 1/z$ rotates the Riemann sphere so that $\infty \mapsto 0$, which allows us to transform everything from infinity to the origin. This leads to the following generalization of Definition 3.2.5, which we state again here as case (i) for convenience of reference. Recall the conventions $1/0 := \infty$ and $1/\infty := 0$.

Definition 3.5.1. A complex function $f : D \subset \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ is said to be *analytic at* $z_0 \in \widehat{\mathbb{C}}$ if z_0 is an inner point of D and one of the following cases occurs:

(i) $z_0 \in \mathbb{C}$, $f(z_0) \in \mathbb{C}$, and there is a neighborhood of z_0 in which f is the sum of a convergent power series centered at z_0,

(ii) $z_0 = \infty$, $f(z_0) \in \mathbb{C}$, and $g(z) := f(1/z)$ is analytic at 0 according to (i),

(iii) $z_0 \in \mathbb{C}$, $f(z_0) = \infty$, and $g(z) := 1/f(z)$ is analytic at z_0 according to (i),

(iv) $z_0 = \infty$, $f(z_0) = \infty$, and $g(z) := 1/f(1/z)$ is analytic at 0 according to (i).

It follows immediately from the definition that $1/f$ is analytic at z_0 if and only if this holds for f. A point $z_0 \in D$ where $f(z_0) = \infty$ is said to be a *pole* of f and its *multiplicity* (or order) is defined as the multiplicity of the zero z_0 of $1/f$ (see Definition 3.3.5).

For example, a function f with finite value $f(\infty)$ is analytic at infinity if and only if there exists a positive r such that

$$f(z) = a_0 + \frac{a_1}{z} + \frac{a_2}{z^2} + \ldots + \frac{a_k}{z^k} + \ldots, \quad |z| > r. \tag{3.63}$$

In the following result continuity has to be understood with respect to the spherical metric.

Theorem 3.5.2. *If $f : D \subset \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ is analytic at $z_0 \in D$ it is continuous at z_0.*

Proof. In case (i) the result follows directly from Theorem 3.2.3. In all other cases f is a composition of the continuous function g with the mappings $z \mapsto 1/z$ and $w \mapsto 1/w$ which are continuous on $\widehat{\mathbb{C}}$. ☐

As earlier, the set of points at which a complex function is analytic is always open in $\widehat{\mathbb{C}}$. In the next step we state the corresponding extension to the definition of analytic functions.

Definition 3.5.3. Let D be an open subset of the Riemann sphere $\widehat{\mathbb{C}}$. A complex function $f : D \to \widehat{\mathbb{C}}$ which is analytic at every point in D in the sense of Definition 3.5.1 is said to be *analytic*.

Classes of Analytic Functions. Traditionally several subclasses of analytic functions are distinguished. We summarize these concepts in the following definition.

Definition 3.5.4. An analytic function is said to be *meromorphic* if all its poles are isolated. An analytic function is called *holomorphic* if it has no poles. For open subsets D of $\widehat{\mathbb{C}}$ we denote by $\mathcal{A}(D)$, $\mathcal{M}(D)$ and $\mathcal{O}(D)$ the sets of all functions $f : D \to \widehat{\mathbb{C}}$ which are analytic, meromorphic and holomorphic in D, respectively.

Saying that all poles of f are isolated means that any point z_0 with $f(z_0) = \infty$ has a neighborhood in which $f(z) \neq \infty$ for $z \neq z_0$. It follows from the identity principle for power series that this condition is automatically satisfied if D contains no disk D_0 on which $f = \infty$. If, on the other hand, such a disk exists, then $f = \infty$ on the connected component of D which contains D_0 (this can be shown by applying Theorem 3.4.1 to $1/f$). If D is a (connected) domain, this implies that $f = \infty$ on all of D. So in this case the sets of analytic functions and meromorphic functions differ only by the constant function $f = \infty$.

We point out that the poles of a meromorphic function f may have accumulation points, but these must lie on the boundary of D and cannot belong to D. Otherwise, by continuity of f, such an accumulation point must be a pole, which would then not be isolated.

The "classical" analytic functions $f : D \subset \mathbb{C} \to \mathbb{C}$, which have been studied in the preceding section, are holomorphic functions. In fact holomorphic functions are precisely those analytic functions which attain only finite values. Note that in the above definition the domain set D of a holomorphic function may contain the point at infinity. This is not completely standard in the literature but it has some advantages.

Local Normal Forms. The following result extends Theorem 3.3.4 to the generalized concept of analytic functions. It can easily be derived by applying Theorem 3.3.4 to the four cases in Definition 3.5.3.

Theorem 3.5.5 (Local Normal Forms)**.** *Let $f : D \subset \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ be analytic and not constant in a neighborhood of $z_0 \in D$. Then there exist a positive integer m and* *an analytic function* $g : D \subset \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ *with* $g(z_0) \neq 0$ *and* $g(z_0) \neq \infty$*, such that one of the following holds for all* $z \in D$*:*

$$z_0 \in \mathbb{C}, \quad f(z_0) \in \mathbb{C}, \quad f(z) = f(z_0) + (z - z_0)^m \, g(z), \tag{3.64}$$

$$z_0 \in \mathbb{C}, \quad f(z_0) = \infty, \quad f(z) = (z - z_0)^{-m} \, g(z), \tag{3.65}$$

$$z_0 = \infty, \quad f(z_0) \in \mathbb{C}, \quad f(z) = f(z_0) + z^{-m} \, g(z), \tag{3.66}$$

$$z_0 = \infty, \quad f(z_0) = \infty, \quad f(z) = z^m \, g(z). \tag{3.67}$$

The integer m *and the function* g *are uniquely determined.*

According to the possible cases we classify the points $z_0 \in D$ as in the following definition. Note that cases (iii) and (iv) give nothing new, they are just included for completeness.

Definition 3.5.6. Let $f : D \subset \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$. A point $z_0 \in D$ is said to be

(i) a *regular point of* f, if (3.64) or (3.66) hold with $f(z_0) \neq 0$ and $m = 1$,

(ii) a *saddle point of* f, if (3.64) or (3.66) hold with $f(z_0) \neq 0$ and $m \geq 2$,

(iii) a *zero of* f *with multiplicity* m, if (3.64) or (3.66) hold with $f(z_0) = 0$,

(iv) a *pole of* f *with multiplicity* m, if (3.65) or (3.67) hold.

The four pictures in Figure 3.33 show enhanced phase portraits of f in a neighborhood of z_0 corresponding (from left to right) to the four cases (i)–(iv). Do pay attention to the opposite orientation of colors for zeros and poles. Note that these pictures look the same whether z_0 is finite or infinite.

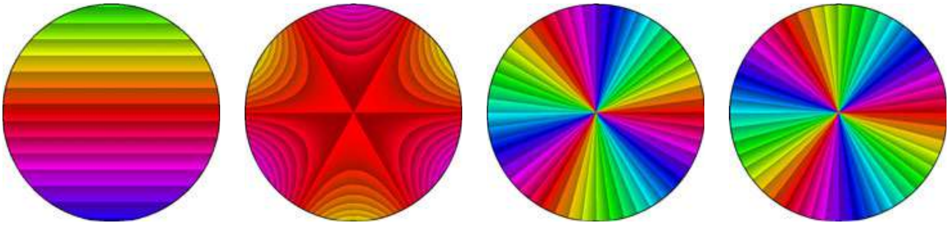


Figure 3.33: Enhanced phase portraits near z_0 corresponding to the normal forms

As in the preceding section, from now on we shall consider only functions which are analytic on domains $D \subset \widehat{\mathbb{C}}$.

Arithmetic Operations. Using the representations in Theorem 3.5.5 and discussing the remaining cases of constant functions separately, it is easy to see that the set $\mathcal{M}(D)$ of meromorphic functions in D is closed with respect to the arithmetic operations, that is, if f, g in $\mathcal{M}(D)$, then also $f \pm g \in \mathcal{M}(D)$, $f\,g \in \mathcal{M}(D)$, and, if g is not the zero function, $f/g \in \mathcal{M}(D)$. This is, by the way, the point where the distinction between meromorphic and analytic functions comes into play. For example, there is no consistent definition of $f + g$ if $f = g = \infty$.

The Fundamental Principles Revisited. Several results of the preceding section remain valid for meromorphic or even analytic functions, like the uniqueness principle.

Theorem 3.5.7 (Identity Theorem, Uniqueness Principle)**.** *Let* $f, g : D \subset \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ *be analytic in a domain* D*. If there exists a sequence* $(z_n) \subset D \setminus \{z_0\}$ *such that* $z_n \to z_0 \in D$ *and* $f(z_n) = g(z_n)$ *for all* $n \in \mathbb{Z}_+$*, then* $f(z) = g(z)$ *for all* $z \in D$*.*

Proof. By continuity we have $f(z_0) = g(z_0)$. If this value is finite, we can proceed as in the proof of Theorem 3.4.1. Otherwise we replace f and g by the analytic functions $1/f$ and $1/g$, respectively. Note that, due to Theorem 3.5.5, Lemma 3.3.6 on isolated zeros of $f - a$ remains valid for analytic functions. ☐

The *open mapping principle* holds for analytic functions $f : D \subset \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ as well. The *maximum principle* does not make sense for analytic functions which attain infinite values, but it can easily be extended to holomorphic functions, that is, analytic functions $f : D \subset \widehat{\mathbb{C}} \to \mathbb{C}$. We leave it to the reader to formulate these results and to think about the necessary modifications in their proofs. An application of the maximum principle is given in the next paragraph.

Meromorphic Functions on the Sphere. It is clear that all rational functions are meromorphic on the Riemann sphere. The next result shows that the converse is also true, which explains why no special name is given to functions which are analytic on $\widehat{\mathbb{C}}$.

Theorem 3.5.8. *If* f *is holomorphic on the Riemann sphere, it is a constant function. If* f *is meromorphic on the Riemann sphere, it is a rational function.*

Proof. 1. Let f be holomorphic on $\widehat{\mathbb{C}}$. Since $\widehat{\mathbb{C}}$ is compact and $|f|$ is continuous, Weierstrass' theorem ensures that $|f|$ attains a (finite) maximum. By the maximum principle, f must be constant.

2. If f is meromorphic on $\widehat{\mathbb{C}}$ its poles are isolated, and since $\widehat{\mathbb{C}}$ is compact, their number must be finite. Repeated application of Theorem 3.5.5, with z_0 running through all *finite* zeros and poles of f, leads to a representation

$$f(z) = (z - z_1)^{m_1} \cdot \ldots \cdot (z - z_k)^{m_k}\, g(z), \quad m_j \in \mathbb{Z}, \tag{3.68}$$

where g is meromorphic on $\widehat{\mathbb{C}}$ and has neither zeros nor poles in $\mathbb{C}$. Depending on whether $g(\infty)$ is finite or infinite, either g or $1/g$ is *holomorphic* on $\widehat{\mathbb{C}}$, and thus must be constant. ☐

The Argument Principle Revisited. A more significant change has to be made in the formulation of the *argument principle* for meromorphic functions. Here we have to count zeros and poles as well. Towards this end we denote by $n(J, f)$ and $p(J, f)$ the numbers of zeros and poles of f in the interior of J, respectively, counted according to their multiplicities.

Theorem 3.5.9 (Argument Principle for Meromorphic Functions)**.** *Let $f : D \to \mathbb{C}$ be meromorphic in the domain $D \subset \mathbb{C}$, let J be a positively oriented Jordan curve with $J \cup \operatorname{int} J \subset D$. If f has neither zeros nor poles on J, then*

$$\operatorname{wind}_J f = n(J, f) - p(J, f).$$

The proof is a simple modification of the proof of Theorem 3.5.9. The only difference is that the integers m_j in the factorization (3.56) (see also (3.68)) are negative if z_j is a pole.

Figure 3.34 illustrates the argument principle for the phase portrait of a typical meromorphic function. Inside the black curve J we find two simple zeros and a pole of multiplicity two. The chromatic number of J is zero. We have oriented the phase flow from the zeros to the poles which is indicated by the arrows in the left picture. The net flux of "phase" through J vanishes.

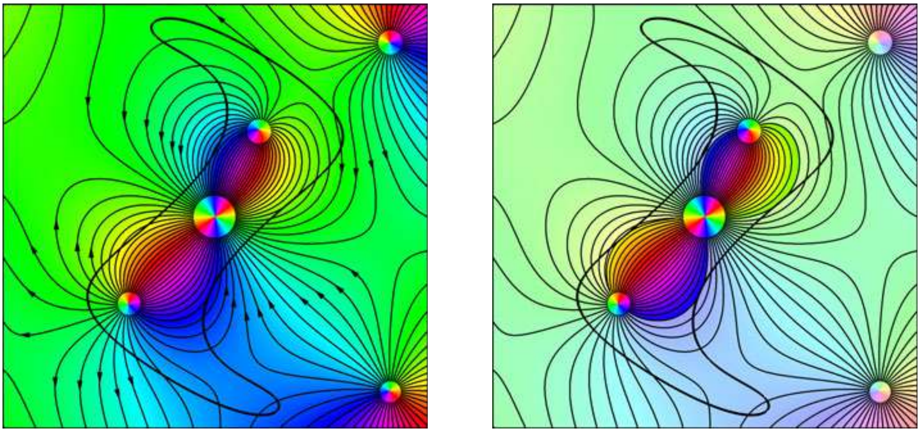


Figure 3.34: Phase flow along the isochromatic lines of a meromorphic function

In contrast to the holomorphic case, not all the "phase" generated by the zeros in J must flow through J. The saturated colors in the picture on the right correspond to that part which is directly absorbed by the pole without leaving the interior of J.

Phase Diagrams. The idea of "phase flow" will be developed further in the second volume of this book. We shall associate with any meromorphic function f in a domain D an *ordinary differential equation* in such a way that the trajectories of its solutions are the isochromatic lines. Endowed with these *orbits*, the phase portrait of a function is converted into a *phase diagram*. The global structure (invariant manifolds, basins of attraction) of phase diagrams will be explored within the framework of *dynamical systems*. Readers who wish to know more right away may consult Wegert [69].

3.6 Analytic Continuation

Typically a function f defined on some set can be extended in many ways to a larger domain. Even if one requires that the extended function inherits additional properties of f, like continuity or differentiability, the extension is usually not unique.

For *analytic* functions we encounter a totally different situation: due to the *uniqueness principle* (Theorem 3.4.1), there is only *one possibility* (if at all) to extend an analytic function to a larger domain. If we know an analytic function on an arbitrarily small open set, say on a disk, it is completely determined. In a sense, an analytic function "is already there" even if we do not yet know it.

This section is devoted to a technique which can be compared with exploring a landscape. At the beginning we stand somewhere (at a point z_0) and can see the function just in a small region (the disk of convergence of its Taylor series at z_0). Moving to the boundary of that region, the horizon changes and we may get access to a new part of the landscape which could not be seen before. Continuing this process, we walk around in the plane and discover more and more of the territory where the function lives.

As in a real journey, it may happen that we cannot go further because some obstacles prevent us from doing so. Sometimes we shall observe strange effects when we go back to some place where we have been before: to our surprise the landscape around might have completely changed since our last visit. And when we go back another time, it may have perhaps returned to its former state – or again changed into something else.

The procedure described for exploring analytic functions is called the *Weierstrass disk chain method.* Its mathematical ingredients are the identity theorem (Theorem 3.4.1), and Weierstrass' rearrangement theorem (Theorem 3.3.1); the first one guarantees that the landscape which we explore is uniquely determined, the second allows us to change the horizon when we are moving around.

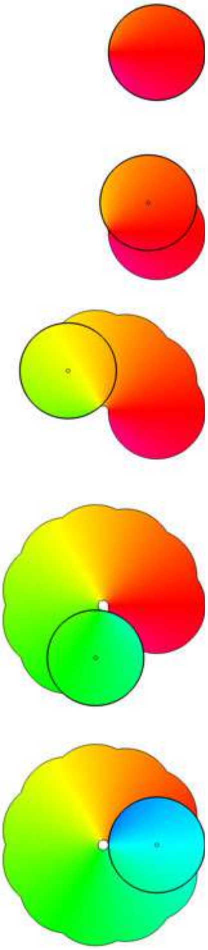


Figure 3.35: Exploring a function

The illustrations in Figure 3.35 demonstrate several steps of an exploration of an analytic function. We start at the center of some disk in the first picture.

Walking along some path in the plane the excursion is documented by taking snapshots that fix the view from certain points t_k along the way. Another hiking tour with the same function and the same starting point, but with "time dependent" sizes of the "horizons" is shown in Figure 3.36.

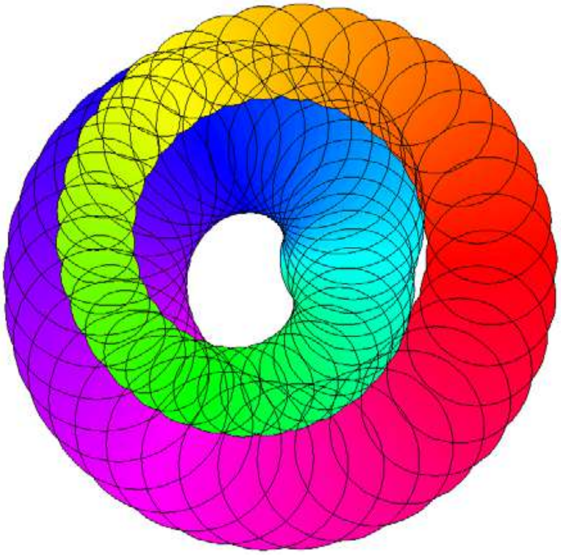


Figure 3.36: An analytic function as a patchwork of function elements

Now let us work out the details. Mainly to avoid notational complications, we restrict ourselves to analytic functions $f : D \subset \mathbb{C} \to \mathbb{C}$.

Direct Analytic Continuation. We begin with a simple result which shows that two analytic functions can be "glued together" when they coincide on the (nonempty) intersection of their domains.

Theorem 3.6.1 (Direct Analytic Continuation, Analytic Extension)**.** *Let the functions $f_1 : D_1 \to \mathbb{C}$ and $f_2 : D_2 \to \mathbb{C}$ be analytic in the domains D_1 and D_2, respectively. Assume that the intersection $D_0 := D_1 \cap D_2$ is not void and $f_1 = f_2$ on D_0. Then there is a unique analytic function f on $D := D_1 \cup D_2$ which coincides with f_1 on D_1, namely*

$$f(z) := \begin{cases} f_1(z) & \text{if} \quad z \in D_1 \\ f_2(z) & \text{if} \quad z \in D_2. \end{cases}$$

Proof. The function f is analytic in D because any point $z \in D$ belongs to D_1 or D_2, so that f coincides with f_1 or f_2 in a neighborhood of z. Since $D_1 \cup D_2$ is a domain (open and connected), and $D_1 \cap D_2 \neq \emptyset$ is open, uniqueness of f follows from the identity theorem (Theorem 3.4.1). $\square$

Under the assumptions of Theorem 3.6.1, the function f is said to be an *analytic continuation* (or *analytic extension*) of f_1 onto D. Interchanging the roles of f_1 and f_2, we see that f is also the (unique) analytic extension of f_2 onto D.

So direct analytic continuation may extend a function to a larger domain, but this says nothing about how to *find* such an extension. The key to a constructive approach is Weierstrass' rearrangement theorem (Theorem 3.3.1) for power series.

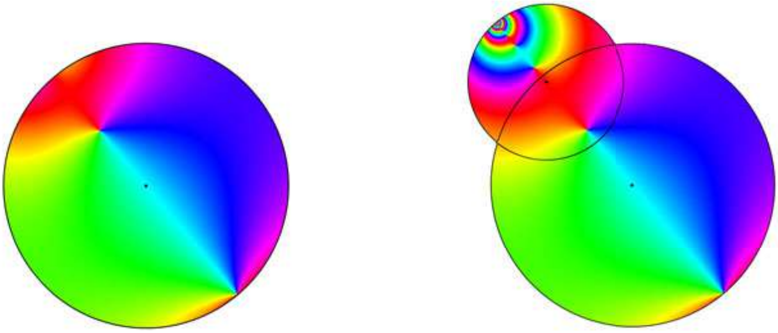


Figure 3.37: A "rearrangement bud"

Analytic Function Elements. Assume that an analytic function f is given as the sum of a power series which has center z_0 and disk of convergence D_0, see Figure 3.37 (left). As we have seen in Example 3.3.1, it can happen that the rearrangement of that power series to a series centered at a point z_1 in D_0 has a disk of convergence D_1 which protrudes out of D_0. Then, by Theorem 3.6.1, f admits an analytic extension to $D_0 \cup D_1$ – the domain of f "sprouted a bud" as shown on the right. In order to explore this further we introduce some notation.

Definition 3.6.2. (i) An *(analytic) function element* is a pair (f, D) consisting of a disk D and an analytic function $f : D \to \mathbb{C}$. The center of the disk D is also referred to as the *center of the function element.*

(ii) If (f_1, D_1) and (f_2, D_2) are two function elements which satisfy the assumptions of Theorem 3.6.1, we say that (f_2, D_2) is the *direct analytic continuation* of (f_1, D_1) (or vice versa) and write $\big(f_1, D_1\big) \text{ ⚭ } \big(f_2, D_2\big)$.

(iii) A finite sequence $\big(f_0, D_0\big)$, $\big(f_1, D_1\big)$, $\ldots$, $\big(f_n, D_n\big)$ of function elements is said to be a *chain*, if any function element (except the first) is the direct analytic continuation of its predecessor,

$$\big(f_0, D_0\big) \text{ ⚭ } \big(f_1, D_1\big) \text{ ⚭ } \quad \ldots \quad \text{ ⚭ } \big(f_n, D_n\big). \tag{3.69}$$

We then call $\big(f_n, D_n\big)$ an *analytic continuation* of $\big(f_0, D_0\big)$ *along the chain.*

(iv) A function element $\big(f_n, D_n\big)$ is an *analytic continuation* of $\big(f_0, D_0\big)$ if a chain of function elements satisfying (3.69) exists. We then write $\big(f_0, D_0\big) \sim \big(f_n, D_n\big)$.

To understand the procedures that follow better it is essential to recognize some subtleties of these definitions. While it is easy to see that $\sim$ is an *equivalence relation*, the relation ⚭ is reflexive and symmetric, but *not transitive*. The latter is clear since $D_1 \cap D_2 \neq \emptyset$ and $D_2 \cap D_3 \neq \emptyset$ does not even imply that $D_1 \cap D_3 \neq \emptyset$. A more interesting situation where transitivity fails is studied next.

Example 3.6.1 (**Analytic Continuation of the Square Root**). We continue to investigate Example 3.3.1. The binomial series

$$f_0(z) = \sum_{n=0}^{\infty} \binom{1/2}{n} (z-1)^n \tag{3.70}$$

has radius of convergence one and thus defines a function element $\big(f_0, D_0\big)$ with $D_0 := \{z \in \mathbb{C} : |z-1| < 1\}$. If z is real and $0 < z < 1$, we have $f_0(z) = \sqrt{z}$.

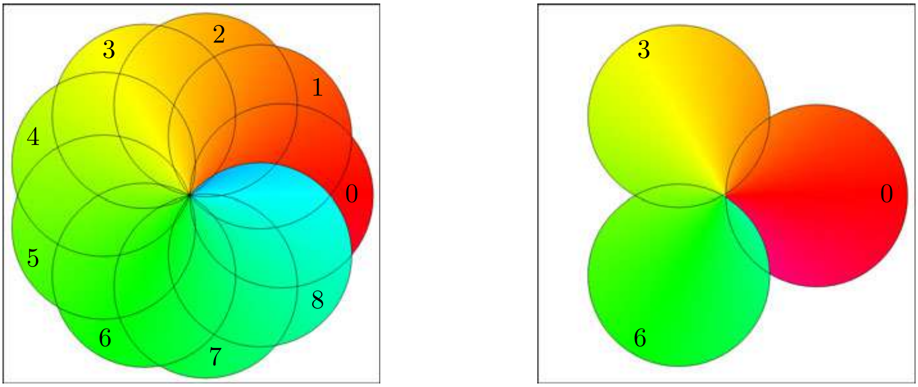


Figure 3.38: Function elements forming a chain (left), ⚭ is not transitive (right)

For $k = 0, 1, \dots, 8$ we denote by $\omega_k = \exp(2k\pi\mathrm{i}/9)$ the 9th roots of unity and let $D_k := \{z \in \mathbb{C} : |z - \omega_k| < 1\}$. All power series

$$f_k(z) := \mathrm{e}^{\mathrm{i}k\pi/18} \sum_{n=0}^{\infty} \mathrm{e}^{-\mathrm{i}k\pi/9} \binom{1/2}{n} (z - \omega_k)^n \tag{3.71}$$

have radius of convergence one and the nine function elements form a chain

$$\big(f_0, D_0\big) \;\text{⚭}\; \big(f_1, D_1\big) \;\text{⚭}\; \dots \;\text{⚭}\; \big(f_8, D_8\big)$$

where neighbors are direct analytic continuations of each other, see Figure 3.38 (left). Consequently any two elements $\big(f_j, D_j\big)$ and $\big(f_k, D_k\big)$ are *analytic continuations* of each other. Moreover, for $k = 1, 2, 3, 4$, the element $\big(f_k, D_k\big)$ is a *direct* analytic continuation of $\big(f_0, D_0\big)$, but not for $k = 5, 6, 7, 8$. Since $\big(f_6, D_6\big)$ is also a direct analytic continuation of $\big(f_3, D_3\big)$, we have, as we see in Figure 3.38 (right)

$$\big(f_0, D_0\big) \;\text{⚭}\; \big(f_3, D_3\big) \;\text{⚭}\; \big(f_6, D_6\big) \not\;\text{⚭}\; \big(f_0, D_0\big),$$

(where $\not\mathrel{\circ\!\!\circ}$ stands for "not $\mathrel{\circ\!\!\circ}$"), which again shows that the relation $\mathrel{\circ\!\!\circ}$ is not transitive.

Lemma 3.6.3. *If* $D_1 \cap D_2 \cap D_3 \neq \emptyset$, $\big(f_1, D_1\big) \mathrel{\circ\!\!\circ} \big(f_2, D_2\big)$ *and* $\big(f_2, D_2\big) \mathrel{\circ\!\!\circ} \big(f_3, D_3\big)$, *then* $\big(f_1, D_1\big) \mathrel{\circ\!\!\circ} \big(f_3, D_3\big)$.

Proof. The functions f_1 and f_3 are analytic in the domain $D_1 \cap D_3$ and coincide (with f_2) on its open subset $D_1 \cap D_2 \cap D_3$. Thus $f_1 = f_3$ on $D_1 \cap D_3$. □

Analytic Continuation Along a Path. In the next step we intend to glue together function elements along a given path. The technical term for this patchwork is *analytic continuation along a path*, the procedure is described in the following definition.

Definition 3.6.4. Let $\gamma : [0, 1] \to \mathbb{C}$ be a path. A chain of function elements

$$\big(f_0, D_0\big) \mathrel{\circ\!\!\circ} \big(f_1, D_1\big) \mathrel{\circ\!\!\circ} \ldots \mathrel{\circ\!\!\circ} \big(f_n, D_n\big) \tag{3.72}$$

is said to be *a chain along* γ, if the chain of disks $(D_0, D_1, \ldots, D_n)$ covers γ in the sense of Definition 2.7.7.

Let $\big(f_0, D_0\big)$ and $\big(f, D\big)$ be function elements with centers at $\gamma(0)$ and $\gamma(1)$, respectively. We say that $\big(f, D\big)$ is an *analytic continuation of* $\big(f_0, D_0\big)$ *along* γ, if there exists a chain of function elements (3.72) along γ such that $(f, D) = (f_n, D_n)$.

Note that the phrase "$(D_0, D_1, \ldots, D_n)$ covers γ" does not simply mean that γ is contained in the union of the disks D_k (for details see Definition 2.7.7).

It is essential that analytic continuation along a path does not depend on the special choice of the chain of function elements. This statement is made precise in the next lemma.

Lemma 3.6.5. *Let* $\big(f_0, D_0\big) \mathrel{\circ\!\!\circ} \ldots \mathrel{\circ\!\!\circ} \big(f_n, D_n\big)$ *and* $\big(g_0, \widetilde{D}_0\big) \mathrel{\circ\!\!\circ} \ldots \mathrel{\circ\!\!\circ} \big(g_m, \widetilde{D}_m\big)$ *be two chains of function elements along a path* γ. *If* $\big(f_0, D_0\big) \mathrel{\circ\!\!\circ} \big(g_0, \widetilde{D}_0\big)$ *then it is also true that* $\big(f_n, D_n\big) \mathrel{\circ\!\!\circ} \big(g_m, \widetilde{D}_m\big)$.

Proof. 1. Let $\gamma : [0, 1] \to \mathbb{C}$. According to Definition 2.7.7, we have partitions

$$0 = t_0 < t_1 < \ldots < t_n = 1, \qquad 0 = s_0 < s_1 < \ldots < s_m = 1,$$

of $[0, 1]$ such that for all $k = 1, \ldots, n$ and $j = 1, \ldots, m$

$$\gamma\big([t_{k-1}, t_k]\big) \subset D_k, \qquad \gamma\big([s_{j-1}, s_j]\big) \subset \widetilde{D}_j.$$

2. Intuitively, the following procedure can be described as a walk along the path γ, where the left foot is only allowed to step on disks D_k, the right foot is restricted to disks $\widetilde{D}_j$, and the function elements $\big(f_k, D_k\big)$ and $\big(f_j, \widetilde{D}_j\big)$ underneath both feet must be direct analytic continuations of each other. We shall show that one can walk step-by-step all the way along γ, following just a simple rule: don't move the foot which is ahead.

We consider the set S of all pairs (k, j) with $1 \le k \le n$ and $1 \le j \le m$ which meet the following condition,

$$[t_{k-1}, t_k] \cap [s_{j-1}, s_j] \neq \emptyset, \qquad \big(f_k, D_k\big) \mathrel{\circ\mkern-7mu\circ} \big(g_j, \widetilde{D}_j\big). \tag{3.73}$$

Since $t_0 = s_0 = \gamma(t_0)$, the intersection of D_1 and $\widetilde{D_1}$ contains a neighborhood of $\gamma(t_0)$, so that (3.73) is satisfied for $k = j = 1$ and S is not empty.

Let (k, j) be an element of S with $k + j < n + m$. Without loss of generality we may suppose that $t_k \le s_j$. Because $[t_{k-1}, t_k] \cap [s_{j-1}, s_j] \neq \emptyset$, this implies $t_k \in [s_{j-1}, s_j]$.

If $k = n$, then $s_j \ge t_n \ge 1$, hence $j = m$ and $k + j = m + n$, a contradiction. Consequently $k + 1 \le n$,

$$t_k \in [t_k, t_{k+1}] \cap [s_{j-1}, s_j] \neq \emptyset,$$

and $\gamma(t_k) \in D_k \cap D_{k+1} \cap \widetilde{D}_j \neq \emptyset$. By Lemma 3.6.3, $\big(f_{k+1}, D_{k+1}\big) \mathrel{\circ\mkern-7mu\circ} \big(f_k, D_k\big)$ and $\big(f_k, D_k\big) \mathrel{\circ\mkern-7mu\circ} \big(\widetilde{g}_j, \widetilde{D}_j\big)$ imply that $\big(f_{k+1}, D_{k+1}\big) \mathrel{\circ\mkern-7mu\circ} \big(\widetilde{g}_j, \widetilde{D}_j\big)$, so that $(k + 1, j) \in S$.

3. Among all elements (k, j) of S, there exists a pair which has the largest sum. From the second step, this pair must be (n, m), because for all other (k, j) in S at least one of the pairs $(k + 1, j)$ or $(k, j + 1)$ also belongs to S. ☐

Example 3.6.2 (**Analytic Continuation of the Square Root II**). We consider the analytic continuation of the square root function, starting with the function element $\big(f_0, D_0\big)$ with $D_0 := \{z \in \mathbb{C} : |z - 1| < 1\}$ and f_0 given by (3.70).

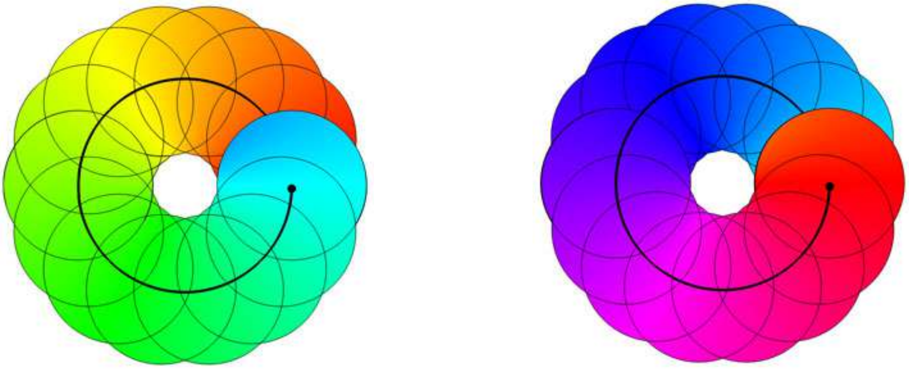


Figure 3.39: Two analytic continuations of the square root along a circular path

Analytic continuation of $\big(f_0, D_0\big)$ along the path $\gamma : [0, 1] \to \mathbb{C}$, $t \mapsto \exp(2\pi i t)$ can be performed using the chain of function elements $\big(f_k, D_k\big)$ with $k = 0, \ldots, 9$ defined in Example 3.6.1. Referring to the definition of f_k in (3.71), we find that $f_9(z) = -f_0(z)$. So, for *any* function element (f, D) which is an analytic continuation of $\big(f_0, D_0\big)$ along γ, the function f coincides with $-f_0$ in a neighborhood of $z_0 = 1$.

Analogously, if (f, D) is an analytic continuation of $\big(-f_0, D_0\big)$ along γ, then $\big(f, D\big) \mathrel{\text{⚭}} \big(f_0, D_0\big)$. In other words: if γ is run through *twice*, we return to the original values of f_0. Figure 3.39 illustrates the result of this procedure, where γ is covered by a chain of 14 function elements defined on disks with radii 0.7.

Function Elements and Germs. Though analytic continuation along a path γ is *essentially* independent of the choice of the function elements which cover γ, these elements are by no means uniquely defined. In fact not even the elements at the endpoints of γ are unique, Lemma 3.6.5 only tells us that the terminal elements of the chain *coincide on some disk* if the initial elements have this property. The redundancy in the process of analytic continuation is sometimes disturbing and makes formulations cumbersome. To eliminate this drawback we utilize the standard technique of *forming classes.*

Definition 3.6.6. Two function elements (f_1, D_1) and (f_2, D_2) centered at z_0 are said to be *equivalent* if $f_1(z) = f_2(z)$ in a neighborhood of z_0. A *germ at* z_0 (or a *germ with center* z_0) is a class of equivalent function elements centered at z_0. The germ which contains a function element (f, D) is denoted by f^*.

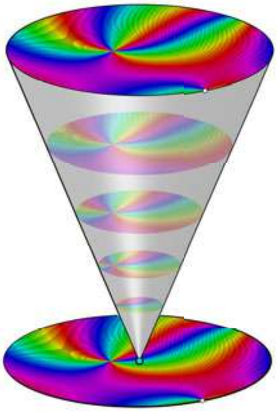


Figure 3.40: Function element and germ

The concept of germs 'compresses' all equivalent function elements centered at z_0 into a single object which 'sits' at the point z_0. Conversely, we may think of the function elements (f, D) in the class f^* as 'growing out' of the germ f^*. Figure 3.40 is an attempt to visualize a germ intuitively as an equivalence class of function elements emerging from their common center.

Depending on the situation, one can choose an appropriate *representative* of a germ f^*, for example a function element (f, D) in f^* with a sufficiently small disk D that fits into some given domain. The *canonical representative* of a germ f^* is that function element (f, D) in f^* which has the disk D of maximal radius (here we allow that $D = \mathbb{C}$). The function element on top of the germ in Figure 3.40 is its canonical representative. It cannot be extended to a larger concentric disk since it has a pole at the marked boundary point.

The *value* $f^*(z_0)$ *of a germ* f^* at z_0 is the value $f(z_0)$ of any function element (f, D) which represents f^*. Note that the value of a germ is only defined at its center.

On the other hand, the germ of a function element (f, D) is *not* determined by the *value* of f at its center z_0 alone, but by the complete list of its Taylor coefficients. So the usual arithmetic operations, inversion and composition of germs are well defined by the corresponding manipulations of power series.

The concept of germs is not restricted to function elements. If the function f is analytic at a point z, it is analytic in a neighborhood of z, and thus it induces a germ at z which we denote by f_z^*.

Analytic Continuation of Germs. Let us now briefly reconsider analytic continuation in the setting of germs.

Definition 3.6.7. We say that a germ f^* at b is an analytic continuation of a germ f_a^* at a along a path γ from a to b, if there exists a chain of function elements $\big(f_0, D_0\big) \text{⚭} \ \ldots \ \text{⚭} \big(f_n, D_n\big)$ along γ such that (f_0, D_0) represents f_a^*, and (f_n, D_n) represents f^*, respectively.

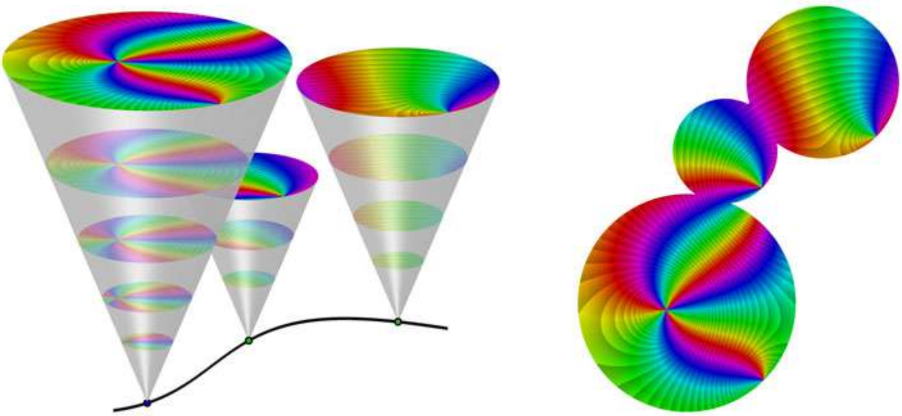


Figure 3.41: Analytic continuation with germs

Whenever an analytic continuation of a germ along a path γ exists, Lemma 3.6.5 tells us that the terminal germ is uniquely determined and does not depend on the specific choice of the function elements along γ. We thus can speak of *the* analytic continuation $f^*(\gamma)$ of a germ f^* along a path γ.

Compared to analytic continuation of function elements, analytic continuation of germs has the further advantage that it does not only yield a unique terminal germ, but also a well-defined *family of germs along* γ. That is, for any t in the parameter interval $[\alpha, \beta]$ of γ, the germ $f^*(\gamma_t)$ is the analytic continuation of f^* along the restriction γ_t of γ to $[\alpha, t]$. Figure 3.41 shows three germs g_1, g_2, g_3 of the family $f^*(\gamma_t)$. The 'view from the top' reveals that the corresponding canonical representatives (f_1, D_1), (f_2, D_2), (f_3, D_3) of the three germs satisfy $(f_1, D_1) \text{⚭} (f_2, D_2)$ and $(f_2, D_2) \text{⚭} (f_3, D_3)$.

The Monodromy Principle. In the next step we study analytic continuation of a germ along *different paths* with the same endpoints. As we have seen in Example 3.6.2, the result may well depend on the selected path γ. On the other hand, there are many situations where analytic continuation is path independent. One such result is stated in the next theorem.

Theorem 3.6.8 (Monodromy Principle I)**.** *Let γ_s, with $s \in [0,1]$, be a family of homotopic paths with fixed endpoints. If the germ f^* admits an analytic continuation $f^*(\gamma_s)$ along any path γ_s, then $f^*(\gamma_0) = f^*(\gamma_1)$.*

Proof. 1. We first prove a *local* result: if $s \in [0,1]$ is fixed, $\sigma \in [0,1]$ and $|\sigma - s|$ is sufficiently small, then $f^*(\gamma_\sigma) = f^*(\gamma_s)$. The proof is a variation of the proof of Lemma 3.6.5.

We fix a representative (f_0, D_0) of f^* and a chain $(f_0, D_0) \circledcirc \ldots \circledcirc (f_n, D_n)$ of function elements covering γ_s, and denote the centers and radii of D_k by z_k and r_k respectively. The corresponding partition $0 = t_0 < t_1 < \ldots < t_n = 1$ of $[0,1]$ satisfies $\gamma_s(t_k) = z_k$ and

$$\gamma_s\big([t_{k-1}, t_k]\big) \subset D_k, \qquad k = 1, \ldots, n.$$

Now we shift the disks D_k so that their centers lie on γ_σ. Using self-explanatory notation, the translated disks can be defined as

$$D_k^\sigma := D_k - \gamma_s(t_k) + \gamma_\sigma(t_k).$$

Invoking a compactness argument, we find that, for all $\sigma \in [0,1]$ with $|\sigma - s|$ sufficiently small, the chain $(D_0^\sigma, D_1^\sigma, \ldots, D_n^\sigma)$ covers γ_σ, and $D_k \cap D_{k+1} \cap D_k^\sigma \neq \emptyset$ for all $k = 0, 1, \ldots, n-1$.

Using the technique in the proof of Lemma 3.6.5, with $\widetilde{D}_j$ replaced by D_j^σ, it now follows that the germs $f^*(\gamma_s)$ and $f^*(\gamma_\sigma)$ have a common representative, that is, $f^*(\gamma_s) = f^*(\gamma_\sigma)$.

2. So far we have shown that the mapping $s \mapsto f^*(\gamma_s)$ is *locally constant*, i.e., for any $s \in [0,1]$ there exists an open interval I_s such that $f^*(\gamma_s) = f^*(\gamma_\sigma)$ for all $\sigma \in I_s \cap [0,1]$. Because $[0,1]$ is compact, it is covered by a finite collection of such (overlapping) intervals, and consequently $s \mapsto f^*(\gamma_s)$ must be constant on all of the interval $[0,1]$. □

Germs Generating Functions. In what follows we fix a germ f^* at z_0 and consider the analytic continuations $f^*(\gamma)$ along *all paths* γ with (fixed) initial point z_0 for which such a continuation exists. The *value* of the germ $f^*(\gamma)$ at the terminal point z of γ is denoted by $f^*(\gamma, z)$. We are interested in conditions which guarantee that $f^*(\gamma, z)$ only depends on the terminal point z and not on the path γ.

Definition 3.6.9. A germ f^* at z_0 is said to have an *unrestricted analytic continuation in a domain* D, if it admits an analytic continuation $f^*(\gamma)$ along *any* path γ in D with initial point z_0. If, moreover, there exists an analytic function $f : D \to \mathbb{C}$ such that $f^*(\gamma, z) = f(z)$ for every $z \in D$, we say that the germ f^* *generates* the analytic function f in D.

The terminology introduced in Definition 3.6.9 will also be applied to the function elements representing f^*.

The following theorem describes some happy situations where *any* germ having an unrestricted analytic continuation generates an analytic function.

Theorem 3.6.10 (Monodromy Principle II)**.** *Let f^* be a germ at $z_0 \in D$ which has an unrestricted analytic continuation in a domain D. Then f^* generates an analytic function in D in each of the following cases:*

(i) *The analytic continuation of f^* along any closed path γ in D is* trivial, *i.e.,*

$$f^*(\gamma, z_0) = f^*(z_0).$$

(ii) *The domain D is simply connected.*

(iii) *The domain D is a punctured simply connected domain (i.e., $D = G \setminus \{a\}$ with a simply connected domain G and $a \in G$) and the analytic continuation of f^* along some closed path γ_0 in D with winding number 1 about a is trivial.*

Proof. (i) Let γ_1 and γ_2 be two paths from z_0 to z in D. The concatenation $\gamma_1 \oplus \gamma_2^-$ of γ_1 with the reverse of γ_2 is a closed path with initial point z_0. Thus, by assumption, $f^*\big(\gamma_1 \oplus \gamma_2^-\big) = f^*$, and consequently

$$f^*\big(\gamma_1\big) = f^*\big(\gamma_1 \oplus \gamma_2^- \oplus \gamma_2\big) = f^*\big(\gamma_2\big).$$

Since the analytic continuation of f^* does not depend on the path from z_0 to z in D, the function $f(z) := f^*(\gamma, z)$ is well defined in D. In order to prove that f is analytic, we choose an arbitrary point z_1 and a path γ_1 from z_0 to z_1 in D. The germ $f_1^* := f^*\big(\gamma_1, z_1\big)$ is represented by a function element $\big(f_1, D_1\big)$. If $z \in D_1$, the path $\gamma := \gamma_1 \oplus [z_1, z]$ connects z_0 with z, and since the segment $[z_1, z]$ lies in D_1, the value of the analytic continuation of f_1^* along $[z_1, z]$ coincides with $f_1(z)$. So we have

$$f(z) = f^*(\gamma, z) = f^*\big(\gamma_1 \oplus [z_1, z], z\big) = f_1^*\big([z_1, z], z\big) = f_1(z), \quad z \in D_1,$$

and hence f is analytic in D_1.

(ii) If D is simply connected, any closed path γ in D with fixed endpoints z_0 is homotopic to the constant path $\gamma_1(t) = z_0$ (see Lemma 2.7.17). By the monodromy theorem (Theorem 3.6.8) the analytic continuation of f^* along γ is trivial, so that (ii) follows from (i).

(iii) If n is a positive integer, we denote by $n \cdot \gamma_0$ the concatenation of n copies of γ_0, we set $-n\,\gamma_0 := n \cdot \gamma_0^-$, and we let $0 \cdot \gamma_0$ stand for the constant path z_0. Then it is clear that $f^*(\gamma_0) = f^*$ implies $f^*(n \cdot \gamma_0) = f^*$ for all $n \in \mathbb{Z}$.

Any closed path γ in D with initial point z_0 and winding number n about a is homotopic with fixed endpoints to the path $n\,\gamma_0$ (see Lemma 2.7.22). The monodromy theorem (Theorem 3.6.8) tells us that $f^*(\gamma) = f^*(n \cdot \gamma_0)$, so that the assumptions of (i) are satisfied. $\square$

Theorem 3.6.10 will play an essential role in the next section and in Section 4.2. Let us now explore what may happen if none of the conditions (i) – (iii) is satisfied. First we consider one of the most prominent functions in the context of analytic continuation.

Example 3.6.3 (The Complex Logarithm). Our starting point is the function element (f_0, D_0) in the disk $D_0 := \{z \in \mathbb{C} : |z-1| < 1\}$ with

$$f_0(z) = \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} \,(z-1)^k = \log|z| + \mathrm{i}\,\mathrm{Arg}\, z, \quad z \in D_0. \tag{3.74}$$

In order to prove that this function element admits an unrestricted analytic continuation in $\mathbb{C} \setminus \{0\}$, we consider any path $\gamma : [0,1] \to \mathbb{C} \setminus \{0\}$ with initial point $z_0 = 1$ and arbitrary terminal point z_1.

In order to construct function elements of an analytic continuation of (f_0, D_0) along γ, we first pick a point $z_t := \gamma(t)$ on γ and denote by

$$D_t := \{z \in \mathbb{C} : |z - z_t| < |z_t|\}, \qquad t \in [0,1],$$

the largest disk around z_t contained in $\mathbb{C} \setminus \{0\}$. To find an appropriate argument of z_t, we denote by $t \mapsto a(t)$ the continuous branch of the argument along γ which is equal to the principal value $\mathrm{Arg}\, 1 = 0$ at its initial point (see Lemma 2.7.19), and set $\arg_\gamma z_t := a(t)$. Finally, we define the function element (f_t, D_t) by

$$f_t(z) := \log|z_t| + \mathrm{i}\, \arg_\gamma z_t + \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k\, z_t^k} \,(z - z_t)^k, \quad z \in D_t. \tag{3.75}$$

The series on the right-hand side results from substituting z by z/z_t in (3.74), so that D_t is indeed its disk of convergence.

Since γ has a positive distance from the origin, there exists a finite subdivision $0 = t_0 < t_1 < \ldots < t_n = 1$ of the interval $[0,1]$ such that the chain of the corresponding disks D_{t_j} covers γ. Thanks to the additional term $\log|z_t| + \mathrm{i}\, \arg_\gamma z_t$ in (3.75), any two neighboring elements $(f_{t_{j-1}}, D_{t_{j-1}})$ and (f_{t_j}, D_{t_j}) are direct analytic continuations of each other, which can be easily verified using (3.74).

The function element (f_1, D_1) at the terminal point z_1 of γ is an analytic continuation of (f_0, D_0) along γ. Following the convention in Definition 3.6.7, we denote the function f_1 by $\log(\gamma, .)$. This function is given by its power series expansion (3.75) with $t = 1$. Using (3.74) we find the explicit representation

$$\log(\gamma, z) = \log|z| + \mathrm{i}\, \arg_\gamma z, \qquad z \in D_1, \tag{3.76}$$

where $\arg_\gamma$ is a distinguished continuous branch of the argument in D_1, namely

$$\arg_\gamma z := \arg_\gamma z_1 + \mathrm{Arg}\,(z/z_1), \quad z \in D_1.$$

Note that the logarithm (3.76) satisfies $\exp\big(\log(\gamma, z)\big) = z$ for all z in D_1.

The Logarithm of a Function. As we have seen in the preceding example, the function $\log z$ cannot be extended from $\mathbb{R}_+$ to an analytic function in the punctured plane $\mathbb{C}\setminus\{0\}$. In this paragraph we show that we can nevertheless define the logarithm of (analytic) functions $f : D \to \mathbb{C}\setminus\{0\}$, provided that D is a *simply* connected domain.

Lemma 3.6.11. *If $f : D \to \mathbb{C}\setminus\{0\}$ is analytic in a simply connected domain D, then there exists an analytic function $g : D \to \mathbb{C}$ such that $f = \exp g$ on D.*

Any function g as in Lemma 3.6.11 is said to be an *analytic branch* of the logarithm of f in D and is denoted by $\log f$.

Proof. Fix a point $z_0 \in D$. If D_0 is a sufficiently small disk centered at z_0, then the image set $f(D_0)$ can be included in some disk G with center $w_0 = f(z_0)$ which does not contain the origin. Denoting by $\log_G$ an analytic branch of the logarithm in G (see the preceding example), and setting $g_0 := \log_G f$ we obtain an analytic branch of the logarithm of f in D_0.

If γ is a path in D with initial point z_0, then $f \circ \gamma$ is a path in $\mathbb{C}\setminus\{0\}$ with initial point w_0. According to the preceding example, the logarithm $\log_G$ has an analytic continuation $\log(f\circ\gamma, .)$ along $f\circ\gamma$. Then the function g defined as the composition

$$g(\gamma, z) := \log\big(f\circ\gamma, f(z)\big), \quad z \in D_\gamma,$$

is an analytic continuation of (g_0, D_0) along γ. Here D_γ is some disk centered at the endpoint of γ. As was shown in Example 3.6.3, we have $\exp\big(g(\gamma, z)\big) = f(z)$ for all $z \in D_\gamma$. Since (g_0, D_0) has an unrestricted analytic continuation in D, and D is simply connected, the monodromy principle (Theorem 3.6.10) yields the existence of the function g. ☐

Remark 3.6.12. It is easy to see that two analytic branches of $\log f$ can only differ by a constant function $2k\pi\mathrm{i}$ with $k \in \mathbb{Z}$. If D does not contain the origin, we can apply Lemma 3.6.11 to $f(z) = z$ and obtain analytic branches of the logarithm function. Any such branch has the form

$$\log : D \to \mathbb{C}, \quad z \mapsto \log|z| + \mathrm{i}\ \arg\ z,$$

where arg is a continuous branch of the argument function in D.

Remark 3.6.13. If $f : D \to \mathbb{C}\setminus\{0\}$ is merely assumed to be *continuous* on the simply connected domain, there exists a continuous function $g : D \to \mathbb{C}$ such that $f(z) = \exp\big(g(z)\big)$ for all $z \in D$. Any such function is called a *continuous branch* of the logarithm of f. In order to construct g, we pick $z_0 \in D$ and fix $\arg f(z_0)$. If γ is a path in D with initial point z_0, there is a continuous branch of $\arg f$ along γ, which we choose to define a continuous branch g_γ of $\log f$ along γ. Then a version of the monodromy principle for continuous functions shows that all the continuations g_γ along different paths fit together to form a continuous function g on D.

Multiple-Valued Functions. In Examples 3.6.2 and 3.6.3 we had obtained different values for the analytic continuation along different paths because the conditions of the monodromy principle were not satisfied. But can something be still retrieved in such a situation?

Even if (f_0, D_0) has an unrestricted analytic continuation in D, any attempt to define an analytic function on D as described in Theorem 3.6.10 must fail if $f(\gamma, z)$ is not solely a function of z but attains different values depending on the choice of the path γ.

Basically we have two options to resolve this dilemma. The (seemingly) simpler solution relies on the concept of *multiple-valued functions*. In this setting the values of a complex function are *sets* of complex numbers.

The analytic continuation of a function element (f_0, D_0) or a germ f_0^* onto a domain D then yields the multiple-valued function

$$F(z) := \big\{ f_0(\gamma, z) : \ \gamma \in \Gamma_D(z) \big\},$$

where $\Gamma_D(z)$ stands for the set of all paths in D with fixed initial point z_0 and (variable) terminal point z. The description of the set on the right-hand side is not very convenient, since it involves the paths γ as parameters. In fact the set $F(z)$ cannot be too large: it is either finite or countable (see Theorem 7.1.2).

Example 3.6.4 (Square Root Function). We continue to investigate Example 3.6.1. The function element (f_0, D_0) with

$$f_0(z) = \sum_{k=0}^{\infty} \binom{1/2}{k} (z-1)^k,$$

and $D_0 := \{z \in \mathbb{C} : |z-1| < 1\}$ is an extension of the real square root function from the interval $(0, 2)$ to D_0. This function element admits an unrestricted analytic continuation to $D := \mathbb{C} \setminus \{0\}$. Let γ be a path from $z_0 = 1$ to z_1 in D, and let (f_1, D_1) be an analytic continuation of (f_0, D_0) along γ. Then the power series expansion of f_1 at z_1 is

$$f_1(z) = r^{1/2} \mathrm{e}^{\mathrm{i}\varphi/2} \sum_{k=0}^{\infty} r^{-k} \mathrm{e}^{-k\mathrm{i}\varphi} \binom{1/2}{k} (z - z_1)^k, \tag{3.77}$$

where $r := |z_1|$ and $\varphi := \arg_\gamma z_1$. Here $\arg_\gamma z_1$ denotes the continuous branch of the argument function along γ with $\arg_\gamma z_0 = 0$ at the initial point $z_0 = 1$ of γ. In particular we have

$$f_1(z_1) = r^{1/2}\, \mathrm{e}^{\mathrm{i}\varphi/2}.$$

The value of φ depends on the path γ, but all possible values differ only by an integer multiple of 2π from the principal value $\operatorname{Arg} z_1$. Consequently, howsoever the path γ has been chosen, $f_1(z_1)$ must always attain one of the two values

$$|z_1|^{1/2} \exp\left(\frac{\mathrm{i}\operatorname{Arg} z_1}{2}\right), \qquad -|z_1|^{1/2} \exp\left(\frac{\mathrm{i}\operatorname{Arg} z_1}{2}\right),$$

where Arg denotes the principal value of the argument. Thus the *multiple-valued square root function* SQRT, obtained by analytic continuation of the function element (3.70) onto $\mathbb{C} \setminus \{0\}$, attains two values at every point $z = r\,\mathrm{e}^{\mathrm{i}\varphi} \neq 0$,

$$\mathrm{SQRT}\,(z) = \left\{ (-1)^k \sqrt{r}\ \mathrm{e}^{\mathrm{i}\varphi/2},\ k = 0, 1 \right\}.$$

The value corresponding to $k = 0$ is called the *principal value* (*principal branch* or *main branch*) of the square root and denoted by $\mathrm{Sqrt}\, z$. Figure 3.42 shows enhanced phase portraits of $\mathrm{Sqrt}\, z$ (left) and $-\mathrm{Sqrt}\, z$ (right).

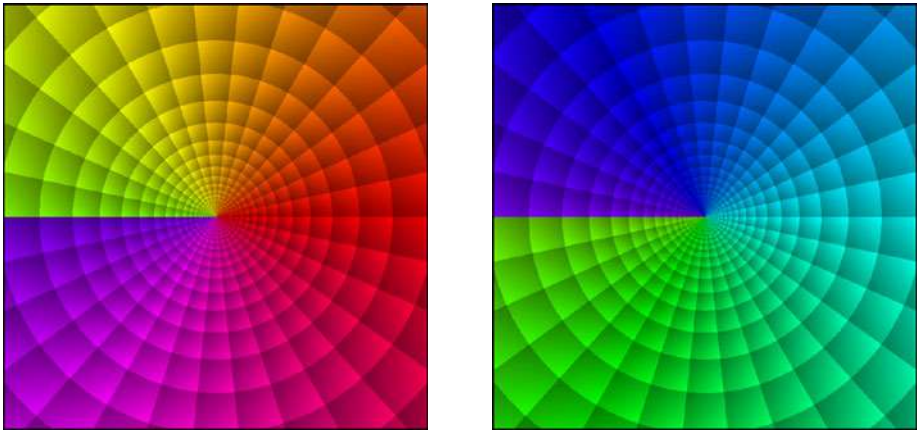


Figure 3.42: Two values of the complex square root function SQRT

Example 3.6.5 (Complex Logarithm Function). As we have seen in Example 3.6.3, the analytic continuation of the function element (f_0, D_0) defined in (3.74) along a path γ in $\mathbb{C} \setminus \{0\}$ from 1 to some point z_1 is given by

$$\log(\gamma, z) = \log|z| + \mathrm{i}\ \arg_\gamma\, z, \qquad |z - z_1| < |z_1|. \tag{3.78}$$

While the real part does not depend on the choice of γ, the imaginary part $\arg_\gamma z$ depends on it and can attain *all* possible values of $\arg z$. Consequently, the *multiple-valued complex logarithm* LOG has a *countable* set of values,

$$\mathrm{LOG}\,(z) = \left\{ \log|z| + \mathrm{i}\,(2k\pi + \mathrm{Arg}\, z) :\ k \in \mathbb{Z} \right\}.$$

Specifying the values of the integer k we get an infinite number of branches of the complex logarithm

$$\mathrm{Log}_k\, z := \log|z| + \mathrm{i}\,(2k\pi + \mathrm{Arg}\, z), \qquad k \in \mathbb{Z}.$$

The function $\mathrm{Log}_0(z)$ is called the *principal value* (or *principal branch*) of the logarithm and also denoted by $\mathrm{Log}(z)$. Figure 3.43 shows three phase portraits of the functions $\operatorname{Log}_k(z)$ in the square $|\operatorname{Re} z| < 2$, $|\operatorname{Im} z| < 2$. All functions are discontinuous along the negative real axis.

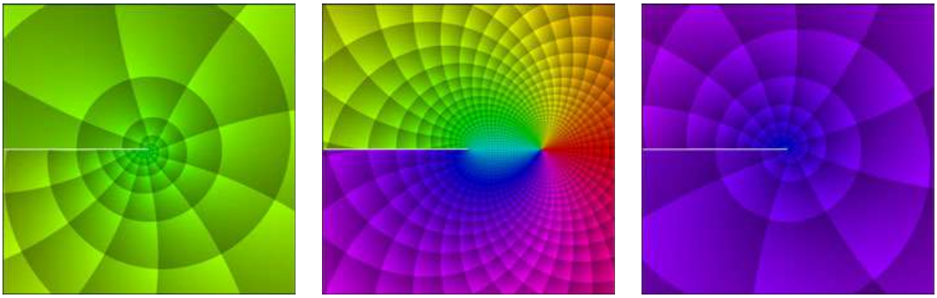


Figure 3.43: Enhanced phase portraits of the branches Log_k for $k = -1, 0, 1$

When working with the complex logarithm one should always be aware of pitfalls. Though $(\exp \circ \operatorname{Log}_k)(z) = z$ for all $k \in \mathbb{Z}$ and $z \in \mathbb{C} \setminus \{0\}$, the concatenation of both functions in reversed order yields a completely different result, which depends on the choice of k and the location of z. This is demonstrated in the phase portraits of Figure 3.44, which visualize the functions $\operatorname{Log}_k \circ \exp$ for $k = -1, 0, 1$ in the square $|\operatorname{Re} z| < 14$, $|\operatorname{Im} z| < 14$. The framed rectangles indicate the regions where $\operatorname{Log}_k\big(\exp(z)\big) = z$.

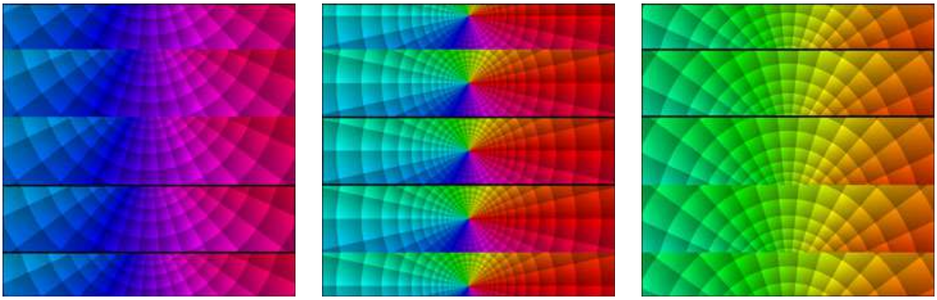


Figure 3.44: Enhanced phase portraits of $\operatorname{Log}_k \circ \exp$ for $k = -1, 0, 1$

Riemann Surfaces. A better alternative to multiple-valued functions rests on a brilliant idea of Bernhard Riemann, who observed that several copies of the complex plane (or of subsets thereof) carrying the branches of a multiple-valued function can be glued together in a clever way to form a new object which is the domain of a *single valued* function. The interested reader may now go directly to Chapter 7 to find out more about this construction.

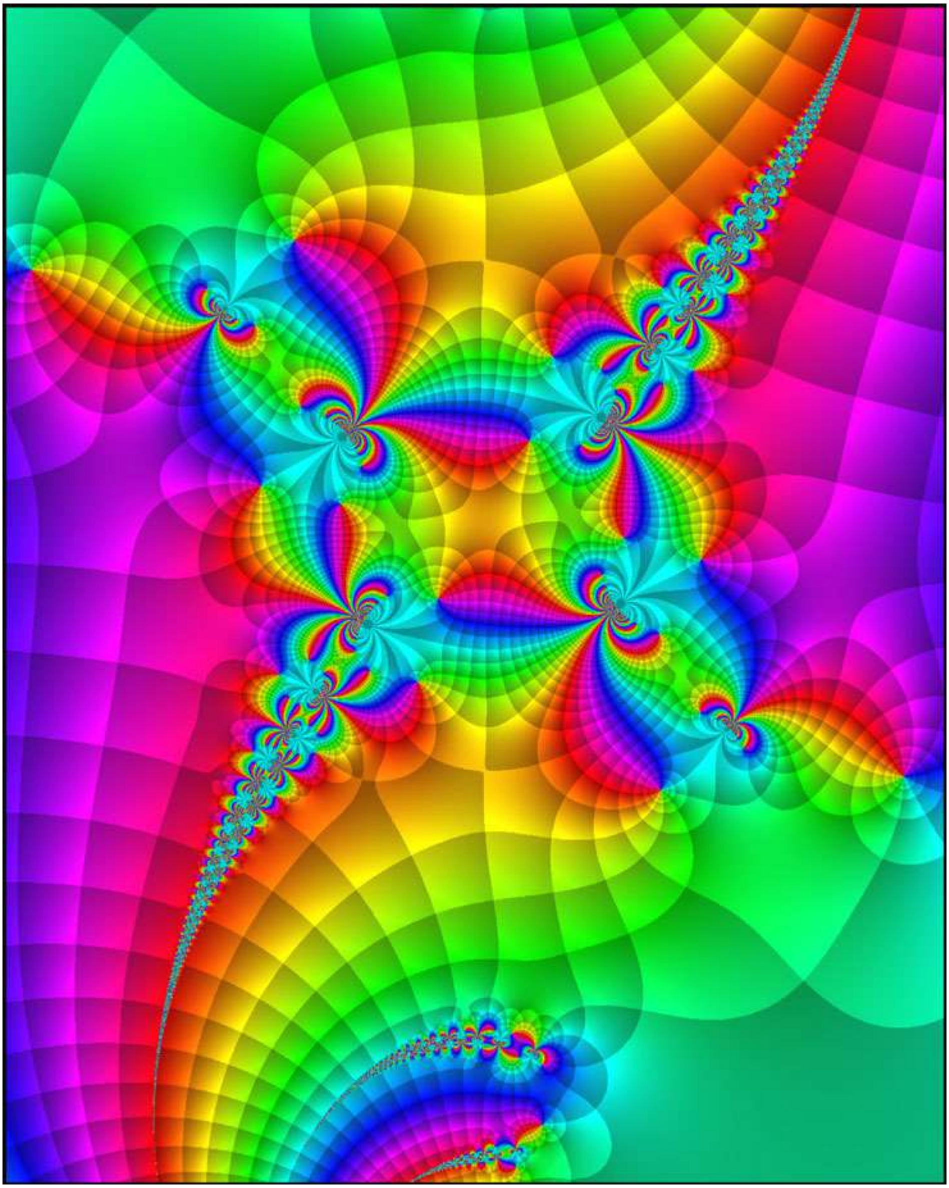


Figure 4.1: Phase portrait of $g \circ f$ with $f(z) := \mathrm{i}z\,\mathrm{e}^{-z/\sin(\mathrm{i}z)} + 1$, $g(z) := f(z) + 2$

Chapter 4

Complex Calculus

The main theme of this chapter is differentiation and integration of complex functions. In Section 4.1 we introduce the concept of the *derivative*, on which *complex calculus* is based. Though there is *formally* no difference between the definitions of differentiability for real and complex functions, the *nature* of these two notions is quite different. This distinction manifests itself in the famous *Cauchy–Riemann equations*, a system of two partial differential equations which couple the real and the imaginary parts of a complex differentiable function.

Complex differentiability is a rather strict requirement and has a number of astounding consequences. For example, in the course of this chapter we shall see that a complex function that is differentiable in an open set, automatically has derivatives of arbitrary order.

As a matter of fact, there is no difference between the classes of analytic and differentiable functions. While it is straightforward to verify that analytic functions are differentiable, the proof of the converse is non-trivial and will accompany us through Sections 4.1 and 4.2. Before the main result (Theorem 4.2.21) can be derived, several other concepts of complex calculus must be discussed.

In particular, in Section 4.2, we shall introduce the *anti-derivative* or *primitive* of a function, which reverses the operation of differentiation, and develop the concept of *path integrals*. We have chosen an approach based on analytic continuation, which allows one to define integrals without making any regularity assumptions on the paths of integration. Later in the same section, we obtain several variants of the *Cauchy integral theorem*, which basically tells us that the integral of an analytic function is invariant with respect to homotopic deformations of the integration path. The highlight is a useful and elegant homology version of *Cauchy's integral formula*.

In Section 4.4, we study functions which are analytic in a domain, with the exception of *isolated singularities*. We investigate the behavior of these functions in a neighborhood of their singularities and state the *residue theorem* which generalizes Cauchy's integral theorem. Among other applications, it proves to be useful

E. Wegert, *Visual Complex Functions: An Introduction with Phase Portraits*,
DOI 10.1007/978-3-0348-0180-5_4, © Springer Basel 2012

for computing various definite integrals which are much harder to evaluate using real variable techniques.

We end the chapter with an excursion into *conjugate harmonic functions* and their applications in plane *potential theory* and *fluid flow*. Though this topic is intimately linked with the Cauchy–Riemann equations, its investigation is deferred to Section 4.6, where we can build on further results derived by then. The reader who is willing to take Theorem 4.2.21 for granted may consult this section directly after Section 4.1.

4.1 Complex Differentiation

In this section we introduce a new concept which is a corner stone of complex function theory: differentiability. Although it is literally defined in the same way as for real functions, there are profound differences in its *consequences* for functions of a real or a complex variable respectively. Since complex differentiability is a rather strong requirement, complex differentiable functions have several peculiar properties, and we shall encounter a number of amazing facts which have no counterparts in real analysis.

The Complex Derivative. In the following we consider exclusively complex functions which are defined on an open subset D of the complex plane and attain values in $\mathbb{C}$.

Definition 4.1.1 (Complex Differentiability)**.** Let D be an open subset of the complex plane. A function $f : D \to \mathbb{C}$ is said to be *differentiable at the point* $z_0 \in D$, if the limit

$$\lim_{z \to z_0} \frac{f(z) - f(z_0)}{z - z_0} \tag{4.1}$$

exists. If this is the case, the limit is called the *derivative* of f at z_0 and denoted by $f'(z_0)$.

The expression $\big(f(z) - f(z_0)\big)/\big(z - z_0\big)$ is designated as the *difference quotient* of f at z_0. Other commonly used notation for the derivative of f at z_0 are

$$\frac{df}{dz}(z_0), \qquad \frac{\partial f}{\partial z}(z_0), \qquad \partial_z f(z_0), \qquad f_z.$$

Linear Approximation. In order to understand the meaning of the derivative, it is convenient to rephrase the definition of differentiability. For fixed $z_0 \in D$ we write

$$f(z) = f(z_0) + f'(z_0)(z - z_0) + r(z), \tag{4.2}$$

where this equation serves as a *definition* of the (remainder) function r, so that

$$\frac{r(z)}{z - z_0} = \frac{f(z) - f(z_0)}{z - z_0} - f'(z_0), \quad z \in D \setminus \{z_0\}.$$

If f is differentiable at z_0 the right-hand side tends to zero as $z \to z_0$, that is the remainder r decays faster than the distance from z to z_0. Using the Landau symbol o (see page 45 for a definition) this can be written as $r(z) = o(z - z_0)$.

It is easy to see that the converse is also true: any function f which is the sum of a linear function $a(z - z_0) + b$ and a function $r(z)$ of order $o(z - z_0)$ is differentiable at z_0 and the coefficient a of the linear term is equal to $f'(z_0)$. In summary, we have the following characterization of complex differentiability:

A complex function is differentiable at z_0 if and only if it can be approximated by a linear function of z, with an error of order $o(z - z_0)$.

Geometric Interpretation. For z sufficiently close to z_0, the error term in equation (4.2) is small, and if $f'(z_0) \neq 0$ it does not play any role in the limit $z \to z_0$. The meaning of this vague statement can be made precise using the Landau symbol $\sim$ (see page 45),

$$f(z) - f(z_0) \sim f'(z_0)(z - z_0).$$

Interpreting this equation geometrically, we see that a differentiable function with $f'(z_0) \neq 0$ acts locally like a “rotostretch” (see page 25): the distance of the image points $f(z)$ and $f(z_0)$ is (approximately) $|f'(z_0)|$ times the distance of the preimage points z and z_0, and the direction of the vector from $f(z_0)$ to $f(z)$ corresponds (approximately) to the direction of the vector from z_0 to z, rotated by an angle of $\arg f'(z_0)$, that is,

$$|f(z) - f(z_0)| \sim |f'(z_0)|\,|z - z_0|, \quad \arg\big(f(z) - f(z_0)\big) \sim \arg(z - z_0) + \arg f'(z_0).$$

This relation also shows that the modulus $|f'(z_0)|$ measures the *amplification* of a small increment $z - z_0$ in the variable z under the action of the function f.

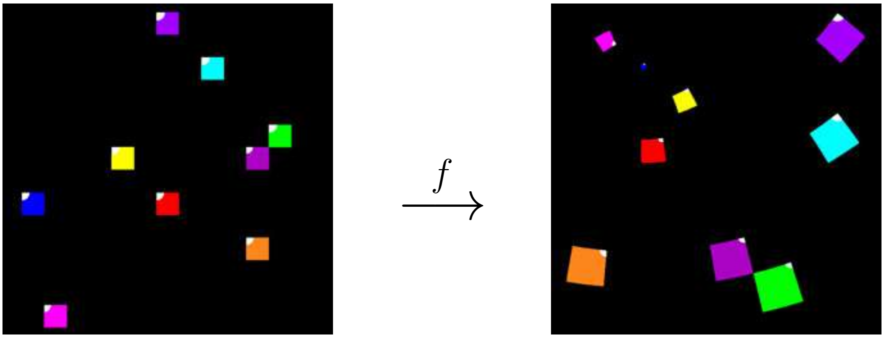


Figure 4.2: A differentiable function mapping small squares to approximate squares

Thus we can say that a differentiable function with $f'(z_0) \neq 0$ acts locally like a *similarity transformation*. It is composed of a dilation with center z_0 and the stretching factor $|f'(z_0)|$, a rotation with center z_0 and rotational angle $\arg f'(z_0)$, followed by a translation which maps z_0 to $f(z_0)$. This is illustrated in Figure 4.2, which shows the images and preimages of several small sets which are approximate squares.

A more detailed analysis of the geometric aspects of differentiability will be given in Chapter 6. Interested readers who are not willing to wait may consult Section 6.1 right now.

Differentiation Rules. Many results about derivatives follow from properties of limits just as in the case of real functions. In order to simplify the formulation of some well known computation rules we first make the following convention.

Saying that a function f is differentiable at z_0 always presupposes that f is defined in a neighborhood of z_0, whether or not this is mentioned at the time.

Theorem 4.1.2 (Differentiation Rules)**.**

(i) *If f and g are differentiable at z and $c \in \mathbb{C}$, then cf, $f+g$, $f-g$, and $f\,g$ are differentiable at z and*

$$\begin{aligned}(cf)'(z) &= c\,f'(z),\\ (f \pm g)'(z) &= f'(z) \pm g'(z),\\ (f\,g)'(z) &= f'(z)\,g(z) + f(z)\,g'(z).\end{aligned}$$

(ii) *If f and g are differentiable at z and $g(z) \neq 0$, then f/g is differentiable at z and*

$$(f/g)'(z) = \frac{f'(z)\,g(z) - f(z)\,g'(z)}{[g(z)]^2}.$$

(iii) *If f is differentiable at z and g is differentiable at $w := f(z)$, then $g \circ f$ is differentiable at z and*

$$(g \circ f)'(z) = g'(w)\,f'(z).$$

Inverse Functions. Recall that a real-valued differentiable function is (locally) invertible in a neighborhood of a point x_0 if $f'(x_0) \neq 0$. In fact this is also true for complex differentiable functions, but there is one technical problem in proving this – the standard real proof utilizes a monotony argument, which we cannot apply in the complex setting. To keep things simple, we first make an additional assumption, which will turn out to be superfluous later, when we have shown that differentiable functions are analytic (see Theorem 4.1.5). We therefore state it as a lemma.

Lemma 4.1.3 (Differentiability of Inverse Functions)**.** *Assume that $f : U \to \mathbb{C}$ is differentiable at z_0 and $f'(z_0) \neq 0$. Let $w_0 := f(z_0)$ and suppose that there is a* continuous *function $g : V \to U$ defined on an open set V containing w_0, such that $f\big(g(w)\big) = w$ for all $w \in V$. Then g is differentiable at w_0 and $g'(w_0) = 1/f'(z_0)$.*

Proof. The function g must be injective on V, since $g(w_1) = g(w_2)$ implies that $w_1 = f(g(w_1)) = f(g(w_2)) = w_2$. Consequently any point $w \in V$ with $w \neq w_0$ can be represented as $w = f(z)$ with $z \in U$ and $z \neq z_0$. Then

$$\frac{g(w) - g(w_0)}{w - w_0} = \frac{g\big(f(z)\big) - g\big(f(z_0)\big)}{f(z) - f(z_0)} = \frac{z - z_0}{f(z) - f(z_0)} \to \frac{1}{f'(z_0)}$$

as $w \to w_0$. Here we have used the fact that g is continuous at w_0, which also guarantees that $z \to z_0$. $\square$

Differentiable Functions. In fact differentiability of a complex function *at a single point* is too weak a property to derive interesting results. However, the situation changes completely if differentiability is assumed *at all points of an open set* D.

We shall not reserve a special name for this class of functions,[1] they are just called *differentiable in* D (or differentiable on D). In fact there is not only no need for another name, it would even be a waste of notation, since we shall see later that the classes of differentiable functions and analytic functions (in the sense of Section 3.4) coincide. But there is still a long way to go before we can prove this.

Polynomials and Rational Functions. Using Theorem 4.1.2 and the trivial facts that $f(z) = 1$ and $g(z) = z$ have derivatives $f'(z) = 0$ and $g'(z) = 1$ respectively, we see that polynomials are differentiable in the entire complex plane and that the derivative of $f(z) = a_0 + a_1\, z + a_2\, z^2 + \ldots + a_n\, z^n$ is

$$f'(z) = a_1 + 2\, a_2\, z + 3\, a_3\, z^2 + \ldots + n\, a_n\, z^{n-1}.$$

A rational function f is differentiable at all points $z \in \mathbb{C}$ where $f(z)$ is finite, and its derivative is a rational function.

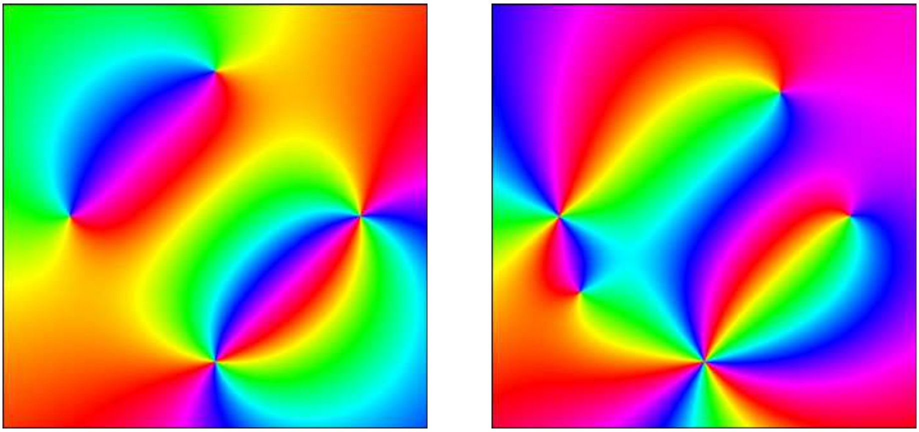


Figure 4.3: Phase portraits of a rational function and its derivative

---

[1] If we had chosen differentiability as the primary concept, we would call these functions "holomorphic" or "analytic"

Example 4.1.1. Figure 4.3 depicts the phase portrait of the rational function

$$f(z) = \frac{(z-1)^2\,(z-\mathrm{i})}{(z+\mathrm{i})^2\,(z+1)}$$

(left) and its derivative f' (right) in the square $|\operatorname{Re} z| < 1.5$, $|\operatorname{Im} z| < 1.5$. Note that f is differentiable in $\mathbb{C} \setminus \{-1, -\mathrm{i}\}$. We see that differentiation decreases the multiplicity of zeros by one, so that simple zero disappears, while the multiplicity of poles is increased by one. It can easily be confirmed that this happens for any rational (and even meromorphic) function. In particular, the derivative of a rational (meromorphic) function can never have a simple pole.

Power Series. The differentiability of sums of power series is a little less obvious. The following result is the easy part of the equivalence between "differentiable" and "analytic" functions. For the converse see Theorem 4.2.21.

Theorem 4.1.4. *If f is analytic in an open set D, then f is differentiable in D.*

Proof. We fix any point z_0 in D. Then there is a disk D_0 with center z_0 where f can be represented by a convergent power series,

$$f(z) = a_0 + a_1\,(z-z_0) + a_2\,(z-z_0)^2 + \ldots = \sum_{k=0}^{\infty} a_k\,(z-z_0)^k, \qquad z \in D_0. \tag{4.3}$$

Then for all $z \in D_0 \setminus \{z_0\}$

$$f(z) - f(z_0) = (z-z_0)\left(a_1 + (z-z_0)\sum_{k=2}^{\infty} a_k\,(z-z_0)^{k-2}\right).$$

The sum of the last series is continuous in D_0, so that dividing by $z - z_0$ and letting $z \to z_0$ gives the desired result. Moreover we see that $f'(z_0) = a_1$. ☐

An explicit formula for the derivative of the sum of a power series follows easily from Weierstrass' rearrangement theorem. If f is given by (4.3), then

$$f'(z) = a_1 + 2\,a_2\,(z-z_0) + 3\,a_3\,(z-z_0)^2 + \ldots + k\,a_k\,(z-z_0)^{k-1} + \ldots, \qquad z \in D_0. \tag{4.4}$$

For a proof we just remark that the right-hand side of (4.4) is just the coefficient b_1 of the rearranged power series (4.3) with center z_1, cf. (3.51) and (3.52).

The formula (4.4) tells us that power series can be differentiated "term-by-term". This can also be shown directly, without recourse to Weierstrass' rearrangement theorem, invoking results on differentiating function series.

Example 4.1.2 (Exponential Function). The exponential function $f(z) = \exp z$ is differentiable in the entire plane and its derivative is

$$f'(z) = \sum_{k=1}^{\infty} \frac{k\,z^{k-1}}{k!} = \sum_{k=0}^{\infty} \frac{z^k}{k!} = \exp z.$$

Similarly we get that the trigonometric functions sin z and cos z are differentiable in the entire complex plane and

$$(\sin z)' = \cos z, \qquad (\cos z)' = -\sin z.$$

Example 4.1.3 (Logarithm Function)**.** For $z \in D := \{z \in \mathbb{C} : |z-1| < 1\}$ the *principal branch* of the complex logarithm is defined by the power series

$$f(z) = \log z := \sum_{k=1}^{\infty} (-1)^{k+1} \frac{(z-1)^k}{k}.$$

Consequently this function is differentiable in D and has the derivative

$$f'(z) = \sum_{k=1}^{\infty} (-1)^{k+1} \frac{k\,(z-1)^{k-1}}{k} = \sum_{k=0}^{\infty} (-1)^k \,(z-1)^k = \frac{1}{1+(z-1)} = \frac{1}{z}.$$

Inverse Functions Revisited. Combining assertion (ii) of Theorem 3.4.11 with Lemma 4.1.3 we obtain the following stronger version of the inverse function theorem for *analytic* functions.

Theorem 4.1.5 (Inverse Function Theorem)**.** *Let f be analytic at z_0 and assume that $f'(z_0) \neq 0$. Then there exists an open disk U centered at z_0 such that f maps U bijectively onto a simply connected domain $V := f(U)$. The inverse mapping $g : V \to U$ is differentiable in V and $g'(f(z)) = 1/f'(z)$ for all $z \in U$.*

The function g is said to be a *local inverse* of f at z_0.

Example 4.1.4. The exponential function $f(z) = \exp z$ is differentiable and its derivative $f'(z) = \exp z$ is different from zero in the entire complex plane. In order to construct a local inverse g of f at a point $z_0 = x_0 + \mathrm{i} y_0$ we set $w_0 := f(z_0)$ and define g in a neighborhood W of w_0 by

$$g(w) = \log|w| + \mathrm{i}\ \arg\, w,$$

where arg denotes a continuous branch of the argument in W, which is chosen such that $\arg\, w_0 = y_0$. Further let U be a disk centered at z_0 such that $V := f(U)$ is contained in W. Then, for all $z \in U$,

$$g\big(f(z)\big) = \log(|\exp z|) + \mathrm{i}\ \arg\,(\exp z) = x + \mathrm{i}\,(y + 2k\pi) = z + 2k\pi\,\mathrm{i}.$$

Since $\arg\, w_0 = y_0$, the integer k is zero for $z = z_0$, and because g is continuous on V, it vanishes for all $z \in U$. Clearly we also have $f\big(g(w)\big) = w$ for all $w \in V$. By Theorem 4.1.5 the function g is differentiable in V and

$$g'(w) = 1/f'\big(g(w)\big) = 1/f\big(g(w)\big) = 1/w.$$

Because different *continuous* branches of the logarithm differ only by a constant function, we always have

$$(\log z)' = \frac{1}{z},$$

no matter which branch has been chosen.

Derivatives of Higher Order. Since (4.4) is a power series representation for the derivative f', Theorem 4.1.4 can be applied recursively. But before stating the result, we introduce some more terminology.

Definition 4.1.6. Assume that the derivative f' of f exists in a neighborhood of z_0. If f' is differentiable at z_0, then $f''(z_0) := (f')'(z_0)$ is called the *second derivative* of f at z_0. The derivatives $f^{(k)}(z_0)$ of higher order k are defined inductively. An *infinitely differentiable* function has derivatives of arbitrary order $k = 1, 2, 3, \ldots$.

It is convenient to consider the function f itself as its zero-th derivative and to set $f^{(0)} := f$. Recursive application of Theorem 4.1.4 yields that analytic functions are infinitely differentiable.

Theorem 4.1.7. *If f is analytic in the open set D, then it is infinitely differentiable in D and all its derivatives are analytic functions. The coefficients a_k of the Taylor series $a_0 + a_1\,(z - z_0) + a_2\,(z - z_0)^2 + \ldots$ of f at a point z_0 in D are given by*

$$a_k = \frac{f^{(k)}(z_0)}{k\,!}, \qquad k = 0, 1, 2, \ldots. \tag{4.5}$$

Real Versus Complex Differentiability. It is now time to explore the difference between real and complex differentiability.

The domain set of a *real* function f is a subset X of the real line. In investigating differentiability of f at an (inner) point $x_0 \in X$, we consider the limit of the difference quotient $\big(f(x) - f(x_0)\big)/(x - x_0)$ as $x \to x_0$. If we think of X as a subset embedded in the complex plane, then the point x can approach x_0 only along the real axis.

In contrast to this, in the definition of complex differentiability, no constraints are imposed with regard to the manner in which the point z approaches z_0,[2] in particular z can tend to z_0 along any direction.

So let us look what happens when z approaches z_0 along straight lines parameterized by $t \mapsto z_0 + t\,\omega$ with $t \in \mathbb{R}_+$ and $|\omega| = 1$. The unimodular parameter ω specifies the direction in which z_0 is approached. In the eight outer windows of Figure 4.4 and Figure 4.5 we have chosen the directions of the eight roots of unity $\omega = \omega_0, \omega_1, \ldots, \omega_7$, the windows are arranged according to the location of ω_k on the unit circle. Each square shows the phase portrait of the corresponding difference quotient

$$f_k(z_0) := \frac{f(z) - f(z_0)}{z - z_0} = \frac{f(z_0 + t\,\omega_k) - f(z_0)}{t\,\omega_k}$$

with $t = 10^{-8}$. Note that the f_k depend on z_0, that is, we visualize the difference quotients of f as functions of z_0 *simultaneously* for all z_0 in the square $|\mathrm{Re}\, z_0| < 1.2$, $|\mathrm{Im}\, z_0| < 1.2$. The central window shows the phase portrait of f.

---

[2] Though the definition of the limit for $z \to z_0$ does not involve points moving towards z_0, we use this picture because it is suggestive.

If f were differentiable at all points of that square, all eight phase portraits would coincide in the limit $t \to 0$, and we would expect that they should look similar if t is sufficiently small. In the first experiment we study the function[3]

$$\begin{aligned} f(x+\mathrm{i}y) = {} & (x^7 - 21x^5y^2 + 35x^3y^4 + x^4 - 7xy^7 - 6x^2y^2 + y^4 + 1) \\ & + \mathrm{i}\,(-7x^6y + 35x^4y^3 + 4x^3y - 21x^2y^5 - 4xy^3 + y^7 + 2y). \end{aligned}$$

The result is visualized in Figure 4.4.

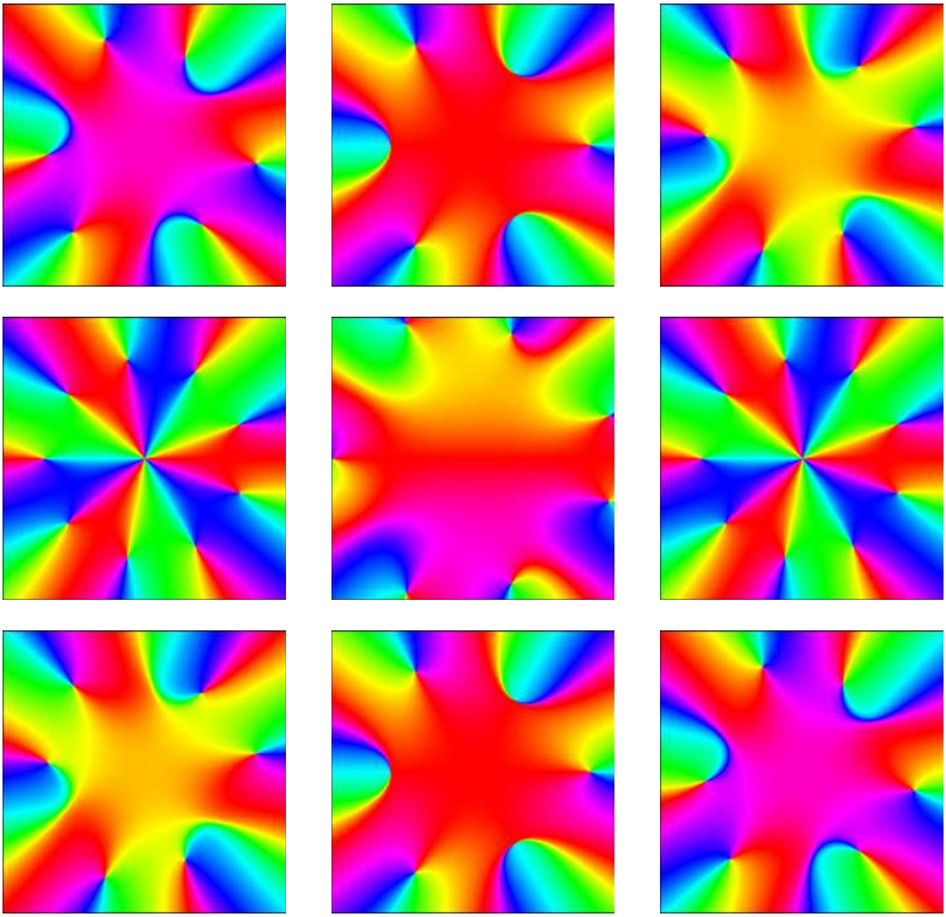


Figure 4.4: Phase portraits of difference quotients of f taken in eight directions

Well, this seems not to have been a good choice with regard to differentiability – these eight pictures look quite incompatible. So let's make a second attempt with the function[4]

---

[3] Such a cryptic formula was chosen in order to make it difficult to guess what will happen, in fact $f(z) = \overline{z^7 - z} + (z^4 + z) + 1$.

$$g(x+\mathrm{i}y) = (x^7 - 21x^5y^2 + 35x^3y^4 + x^4 - 7xy^7 - 6x^2y^2 + y^4 + 1) \\ + \mathrm{i}\,(7x^6y - 35x^4y^3 - 4x^3y + 21x^2y^5 - 4xy^3 - y^7).$$

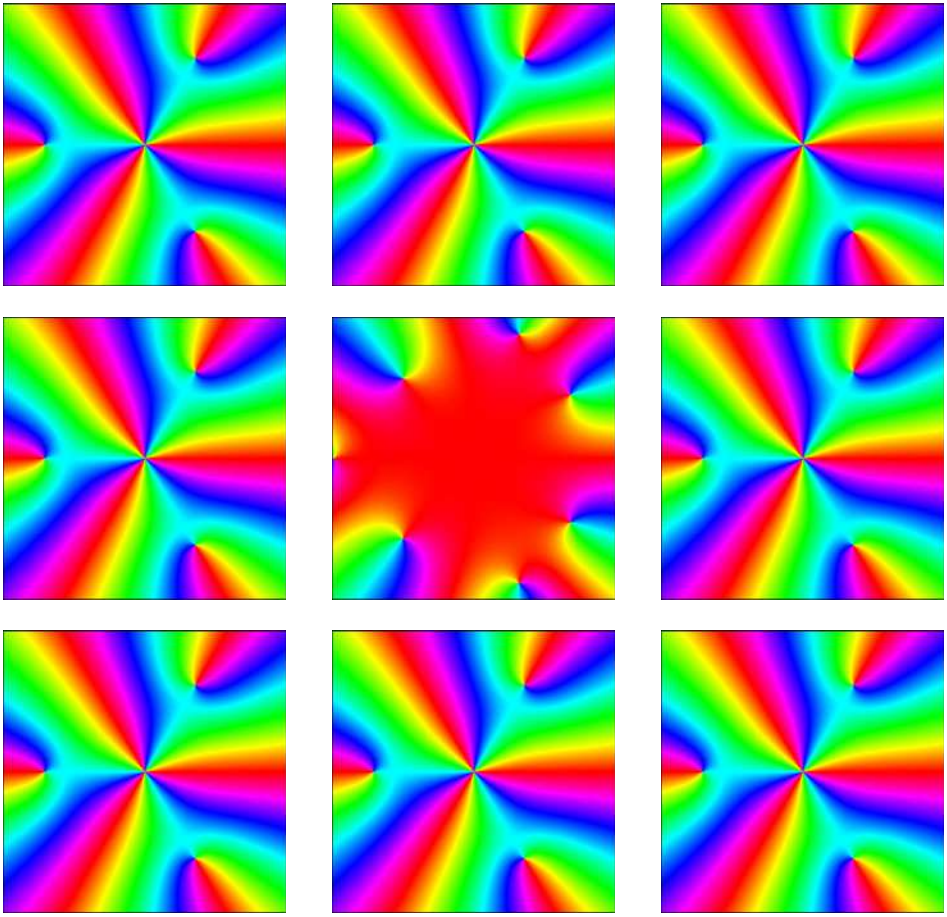


Figure 4.5: Phase portraits of difference quotients of g taken in eight directions

Figure 4.5 shows the result (with g depicted in the middle). This time we see no difference in the eight phase portraits, so that an optimist may hope that g is differentiable at all points of the square depicted. Indeed it is, but how can we check this? To this end let us examine next the limit in (4.1) in more detail.

---

[4] A simpler formula for this function is $g(z) = z^7 + z^4 + 1$.

The Cauchy-Riemann equations. Assume that f is differentiable at $z_0 = x_0 + \mathrm{i}\, y_0$. Emphasizing the real and imaginary parts of z and f, we represent f in the form

$$f(x + \mathrm{i}\, y) = u(x, y) + \mathrm{i}\, v(x, y)$$

where u and v are real-valued functions of two real variables x and y. If z converges to z_0 along a path parallel to the real axis, we have $z = x + \mathrm{i} y_0$ with $x \to x_0$. Since the limit of the difference quotient is equal to $f'(z_0)$, we find

$$\begin{aligned} f'(z_0) &= \lim_{x \to x_0} \frac{f(x + \mathrm{i}\, y_0) - f(x_0 + \mathrm{i}\, y_0)}{x - x_0} \\ &= \lim_{x \to x_0} \frac{u(x, y_0) - u(x_0, y_0)}{x - x_0} + \mathrm{i} \lim_{x \to x_0} \frac{v(x, y_0) - v(x_0, y_0)}{x - x_0} \\ &= \frac{\partial u}{\partial x}(x_0, y_0) + \mathrm{i}\, \frac{\partial v}{\partial x}(x_0, y_0), \end{aligned} \tag{4.6}$$

where $\partial u/\partial x$ and $\partial v/\partial x$ denote the *partial derivatives* of u and v with respect to x. Note that the limits in the second line of (4.6) are just the definitions $\partial u/\partial x$ and $\partial v/\partial x$, and that the limits must exist because the left-hand side has a limit. Similarly, if z approaches z_0 along a line parallel to the imaginary axis, we have $z = x_0 + \mathrm{i}\, y$, and for $y \to y_0$ we get

$$\begin{aligned} f'(z_0) &= \lim_{y \to y_0} \frac{f(x_0 + \mathrm{i}\, y) - f(x_0 + \mathrm{i}\, y)}{\mathrm{i}\,(y - y_0)} \\ &= -\mathrm{i} \lim_{y \to y_0} \frac{u(x_0, y) - u(x_0, y_0)}{y - y_0} + \lim_{y \to y_0} \frac{v(x_0, y) - v(x_0, y_0)}{y - y_0} \\ &= \frac{\partial v}{\partial y}(x_0, y_0) - \mathrm{i}\, \frac{\partial u}{\partial y}(x_0, y_0). \end{aligned} \tag{4.7}$$

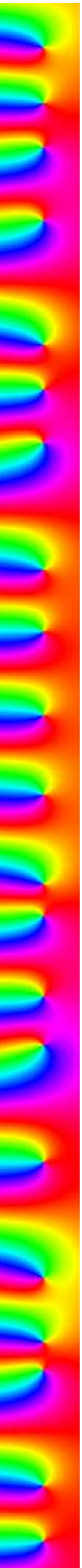
So we have two representations for the derivative,

$$f'(z_0) = \frac{\partial u}{\partial x}(x_0, y_0) + \mathrm{i}\, \frac{\partial v}{\partial x}(x_0, y_0) = \frac{\partial v}{\partial y}(x_0, y_0) - \mathrm{i}\, \frac{\partial u}{\partial y}(x_0, y_0). \tag{4.8}$$

Comparing the real and imaginary parts leads to the conclusion that differentiability of a complex function implies certain relations between the partial derivatives of its real and imaginary parts. Since we are usually not interested in differentiability at a single point, we summarize our investigations as follows.

Theorem 4.1.8 (Cauchy-Riemann Equations)**.** *If $f(x + \mathrm{i}y) = u(x, y) + \mathrm{i}\, v(x, y)$ is differentiable in the open set D, then the partial derivatives of u and v with respect to x and y exist in D and satisfy the* Cauchy-Riemann equations

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}, \qquad \frac{\partial v}{\partial x} = -\frac{\partial u}{\partial y}. \tag{4.9}$$



Example 4.1.5. The function $f(z) = \overline{z}$ is nowhere differentiable, because we have $u(x,y) = x$, $v(x,y) = -y$, so that

$$\frac{\partial u}{\partial x} = 1 \neq -1 = \frac{\partial v}{\partial y}.$$

Necessary and Sufficient Conditions. The Cauchy-Riemann equations (4.9) form a system of *partial differential equations.* They give *necessary* conditions for a function to be (complex) differentiable. The question whether these equations are also *sufficient* is more delicate. The validity of the Cauchy-Riemann equations *at a single point* z_0 does *not guarantee* that f is differentiable at z_0. Consider, for example, the function f defined by $f(x+\mathrm{i}y) = 0$ if $xy = 0$ and $f(x+\mathrm{i}y) = 1$ otherwise. Then all four partial derivatives at $z_0 = 0$ are zero, but f is not even continuous at this point.

However, adding supplementary information can improve the situation. For example, we can make the *a-priori* assumption that f is (Fréchet) differentiable as a function of the *real* variables x and y. The following definition explains what this means.

Definition 4.1.9. A complex function $f : D \to \mathbb{C}$ is said to be $\mathbb{R}$-*differentiable* at a point z_0 in the open set D if there are constants $A, B \in \mathbb{C}$ such that f can be represented as

$$f(z) = f(z_0) + A\,(x - x_0) + B\,(y - y_0) + r(z),$$

where $z = x + \mathrm{i}y$, $z_0 = x_0 + \mathrm{i}y_0$, and the remainder r satisfies $r(z) = o(|z - z_0|)$, i.e. $r(z)/|z - z_0| \to 0$ as $z \to z_0$.

It is easy to see that a function f which is $\mathbb{R}$-differentiable at z_0 has partial derivatives at z_0,

$$\frac{\partial f}{\partial x}(z_0) := \frac{\partial u}{\partial x}(x_0, y_0) + \mathrm{i}\,\frac{\partial v}{\partial x}(x_0, y_0), \qquad \frac{\partial f}{\partial y}(z_0) := \frac{\partial u}{\partial y}(x_0, y_0) + \mathrm{i}\,\frac{\partial v}{\partial y}(x_0, y_0),$$

and that these derivatives coincide with A and B, respectively. Consequently we have

$$f(z) = f(z_0) + \frac{\partial f}{\partial x}(z_0)(x - x_0) + \frac{\partial f}{\partial y}(z_0)(y - y_0) + o(|z - z_0|). \tag{4.10}$$

Conversely, a standard result in real analysis tells us that any function which has *continuous* partial derivatives *in a neighborhood* of z_0 is $\mathbb{R}$-differentiable at z_0.

The Cauchy-Riemann Operator. In order to rewrite the representation (4.10) in a form which is more convenient for our purposes we introduce the notation

$$\frac{\partial f}{\partial z} := \frac{1}{2}\left(\frac{\partial f}{\partial x} - \mathrm{i}\,\frac{\partial f}{\partial y}\right), \qquad \frac{\partial f}{\partial \overline{z}} := \frac{1}{2}\left(\frac{\partial f}{\partial x} + \mathrm{i}\,\frac{\partial f}{\partial y}\right), \tag{4.11}$$

which converts (4.10) into

$$\frac{f(z)-f(z_0)}{z-z_0}=\frac{\partial f}{\partial z}(z_0)+\frac{\overline{z-z_0}}{z-z_0}\,\frac{\partial f}{\partial \overline{z}}(z_0)+o(1). \tag{4.12}$$

Since the limit of $(\overline{z-z_0})/(z-z_0)$ for $z\to z_0$ does not exist, it is now clear that f is complex differentiable at z_0 if and only if

$$\frac{\partial f}{\partial \overline{z}}(z_0)=0, \tag{4.13}$$

in which case we have

$$f'(z_0)=\frac{\partial f}{\partial z}(z_0).$$

Separating the real and imaginary parts in (4.13), we see that this condition is just a reformulation of the Cauchy-Riemann equations. The differential operator $\partial f/\partial\overline{z}$ is therefore called the *Cauchy-Riemann operator*. Both operators introduced in (4.11) are also known as *Wirtinger derivatives*.

Summarizing the above considerations we obtain the following result.

Theorem 4.1.10. *A function $f : D\to\mathbb{C}$ which is $\mathbb{R}$-differentiable at $z_0\in D$ is (complex) differentiable at z_0 if and only if it satisfies the Cauchy-Riemann equation at z_0.*

The Complex Differential. The linear part of the increment $f(z)-f(z_0)$ in (4.12) is said to be the (complex) *differential* of f at z_0 and denoted by df. Defining $dz:=z-z_0$ and $d\overline{z}:=\overline{z-z_0}$ we have

$$df=\frac{\partial f}{\partial z}(z_0)\,dz+\frac{\partial f}{\partial \overline{z}}(z_0)\,d\overline{z}. \tag{4.14}$$

So one can say that the function f is complex differentiable at z_0 if and only if its differential does not depend on $d\overline{z}$.

Polar Coordinates. The Cauchy-Riemann equations can also be expressed in terms of polar coordinates r and φ. If $z\neq 0$, we can choose a continuous branch of the argument φ in the neighborhood of z. Then x and y are differentiable functions of r and φ,

$$x=r\cos\varphi,\qquad y=r\sin\varphi.$$

Applying the chain rule for partial derivatives of real-valued functions the Cauchy-Riemann equations at points different from the origin are transformed to

$$\frac{\partial u}{\partial r}=\frac{1}{r}\frac{\partial v}{\partial \varphi},\qquad \frac{\partial v}{\partial r}=-\frac{1}{r}\frac{\partial u}{\partial \varphi}. \tag{4.15}$$

After some standard calculations we get a representation for the derivative of f,

$$f'(z)=\frac{\overline{z}}{|z|}\left(\frac{\partial u}{\partial r}(z)+\mathrm{i}\,\frac{\partial v}{\partial r}(z)\right). \tag{4.16}$$

Example 4.1.6. Let D be a simply connected domain D which does not contain the origin, and let f be an analytic branch of the logarithm on D,

$$f(z) = \log|z| + \mathrm{i}\,\arg\, z = \log r + \mathrm{i}\varphi$$

(see Remark 3.6.12). We already know that this function is analytic in D and hence differentiable. Verification of the Cauchy–Riemann equations is much easier in polar coordinates,

$$\frac{\partial u}{\partial r} = \frac{1}{r} = \frac{1}{r}\,\frac{\partial v}{\partial \varphi} = 1, \qquad \frac{\partial v}{\partial r} = 0 = -\frac{1}{r}\,\frac{\partial u}{\partial \varphi} = 0,$$

and from (4.16) we again obtain that $f'(z) = \overline{z}/|z|^2 = 1/z$ (see also Example 4.1.4).

The Logarithmic Derivative. The *logarithmic derivative* of a differentiable function is the quotient f'/f. It is well defined at all points z where $f(z) \neq 0$, but we shall see later that it has a natural extension to the zeros of f as well.

The name "logarithmic derivative" is motivated by the following observation. If $w_0 := f(z) \neq 0$ for some z_0, then we can choose an analytic branch of the logarithm in a disk centered at w_0. Since f is continuous, $g(z) := \log f(z)$ makes sense in a neighborhood of z_0. By the chain rule, the composition g is differentiable and $g'(z) = f'(z)/f(z)$. This result is independent of the branch chosen, and finally one does not need the logarithm at all to define the logarithmic derivative.

Error Analysis. The logarithmic derivative is of great importance and has many applications. Here is one application in error analysis: if we think of $z - z_0$ as an error in the variable z (the deviation of z from z_0), then $|f'(z_0)|$ measures its amplification under the action of f. Likewise, the modulus of the logarithmic derivative $|f'(z_0)/f(z_0)|$ measures this effect for the *relative error* $\big(f(z) - f(z_0)\big)/f(z_0)$.

Isochromatic Lines and Contour Lines. More interesting for our purposes is an application of the logarithmic derivative to some aspect in the interpretation of phase portraits.

We consider a function f which is *analytic* at a point z_0, and assume that z_0 is a *regular point* of f, that is, $f(z_0) \neq 0$ and $f'(z_0) \neq 0$ (see Definition 3.5.6). Then there exists a disk D centered at z_0 where f is analytic and all points are regular. By Lemma 3.6.11 the function f has an analytic logarithm $\log f$ in D,

$$g := \log f = \log|f| + \mathrm{i}\,\arg\, f = u + \mathrm{i}\,v, \tag{4.17}$$

where arg stands for a continuous branch of the argument in D. Since $\log f$ is analytic in D, the real functions $u = u(x, y)$ and $v = v(x, y)$ are infinitely differentiable with respect to x and y. By the Cauchy–Riemann equations (4.9) we have

$$g' = \frac{\partial u}{\partial x} + \mathrm{i}\,\frac{\partial v}{\partial x} = \frac{\partial u}{\partial x} - \mathrm{i}\,\frac{\partial u}{\partial y} = \frac{\partial v}{\partial y} + \mathrm{i}\,\frac{\partial v}{\partial x}.$$

Because $g' = f'/f$ has no zeros in D, the *gradients* of u and v do not vanish in D,

$$\left(\frac{\partial u}{\partial x}(z), \frac{\partial u}{\partial y}(z)\right) \neq (0,0), \qquad \left(\frac{\partial v}{\partial x}(z), \frac{\partial v}{\partial y}(z)\right) \neq (0,0), \qquad z \in D.$$

Using the implicit function theorem in $\mathbb{R}^2$ we conclude that the *contour lines*

$$\{x + \mathrm{i}\, y \in D : \ u(x,y) = \text{const}\}, \qquad \{x + \mathrm{i}\, y \in D : \ v(x,y) = \text{const}\}$$

of u and v are smooth curves. Equation (4.17) tells us that the contour lines of u are exactly the lines on which $|f| = \text{const}$, while the contour lines of v are the *isochromatic lines* on which the argument (and hence the phase) of f is constant. Since the gradient of a function is orthogonal to its contour lines, we see that the isochromatic lines are the lines of *steepest descent* (or ascent) of $|f|$.

In which direction along an isochromatic line a function increases or decreases can be read off from the coloring in its neighborhood. For example, according to the color scheme chosen, we have red on the right and green on the left when walking on a yellow line in ascending direction. This is demonstrated in Figure 4.6 which shows contour lines in the analytic landscape and the phase portrait of the rational function $f(z) = (z-1)/(z^2+z+1)$.

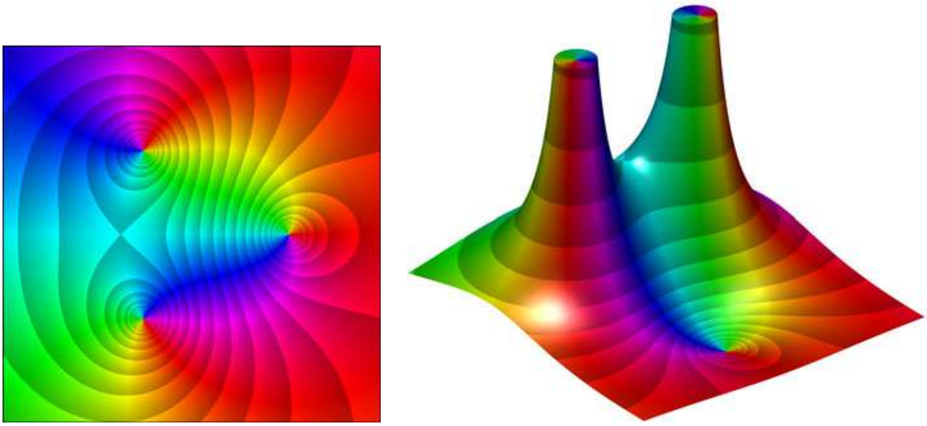


Figure 4.6: Contour lines in the phase portrait and on the analytic landscape

The Growth Rate of a Function. In fact the phase plot does not only tell us the *direction* of the gradient of $|f|$, but also the *growth rate* of $|f|$, which is $|f'|/|f|$, by definition. Roughly speaking, smaller the distance between neighboring isochromatic lines, faster the function grows.

To go a little beyond this qualitative result, we denote by s the unit vector parallel to the gradient of $|f|$ and set $n := \mathrm{i}s$. A short calculation using the

Cauchy-Riemann equations for $\log f$ yields that the directional derivatives of u and v satisfy

$$\frac{\partial u}{\partial s} = \frac{\partial v}{\partial n} > 0, \qquad \frac{\partial u}{\partial n} = -\frac{\partial v}{\partial s} = 0$$

$$|(\log f)'|^2 = |g'|^2 = \left(\frac{\partial v}{\partial n}\right)^2 + \left(\frac{\partial v}{\partial s}\right)^2 = \left(\frac{\partial v}{\partial n}\right)^2$$

at all points $z = x + \mathrm{i}y$ in D. Consequently $\partial v/\partial n$ is just the modulus of the logarithmic derivative. On the other hand, $\partial v/\partial n$ measures the *density of the isochromatic lines*, so that we can visually estimate the growth of $\log|f|$ along these lines from their density in the phase portrait. Because the phase portrait delivers no information on the absolute value of f, it is clear that we can only determine its *relative growth.*

Ideal examples for calibration are the exponential functions $f(z) = \exp(az)$ with a positive parameter a. These functions grow exponentially with the factor a in the direction of the positive real axis. All isochromatic lines are parallel and the distance between two such lines having the same color is $2\pi/a$. Since the logarithmic derivative of f equals a, we *define* the density of the isochromatic lines by $|f'|/(2\pi|f|)$.

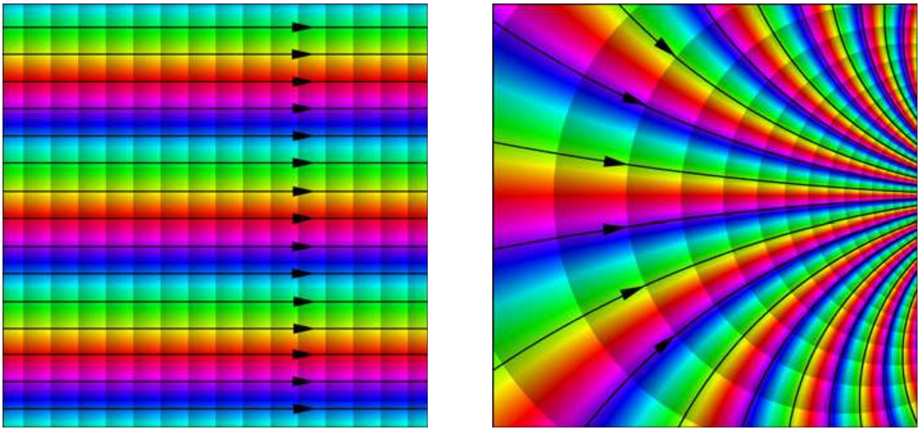


Figure 4.7: Exploring the relative growth of a function in its phase portrait

Figure 4.7 depicts two phase portraits with enhanced contour lines of $|f|$ generated by shading. On each line the absolute value of f is constant, and from one line to the next (in the direction indicated by the arrows) it grows by the same factor. The picture on the left shows an exponential function, which has the same growth rate everywhere. The function on the right grows fastest near the middle of the right edge of the square.

Saddle Points. The above considerations make no sense if z_0 is a zero of f or f', but the structure of the phase portrait in a neighborhood of z_0 can be explored using Theorem 3.3.4. If z_0 is a zero of f with multiplicity m, then there exists an analytic function g with $g(z_0) \neq 0$ such that $f(z) = (z - z_0)^m g(z)$. This happens if and only if

$$f(z_0) = f'(z_0) = \ldots = f^{m-1}(z_0) = 0, \qquad f^m(z_0) \neq 0. \tag{4.18}$$

In this case exactly m isochromatic lines of any specific color emerge at z_0 (see Figure 3.23). Note that, according to convention, the point z_0 itself is colored black, so that it does not belong to any isochromatic line.

If $f(z_0) \neq 0$ and $f'(z_0) = 0$, then, by Theorem 3.3.4, f can be represented as $f(z) = f(z_0) + (z - z_0)^{m+1} g(z)$, with $m \geq 1$ and $g(z_0) \neq 0$. By Definition 3.5.6 such points are denoted as *saddle points*, and we refer to m as the *order* of the saddle point. So a saddle point of order m is characterized by the conditions

$$f(z_0) \neq 0, \quad f'(z_0) = \ldots = f^m(z_0) = 0, \quad f^{m+1}(z_0) \neq 0. \tag{4.19}$$

In a neighborhood of a saddle point z_0 of order m, the set of all points which carry the same color as z_0 forms a star which consists of the point z_0 and $2m + 2$ smooth rays emerging at that point. When z moves away from z_0, then the rays on which $|f|$ increases and the rays on which $|f|$ decreases are alternating. In the analytic landscape, saddle points look like what their name suggests. They correspond exactly to those points where the tangent plane to the graph of $|f|$ is "horizontal", i.e., parallel to the z-plane.

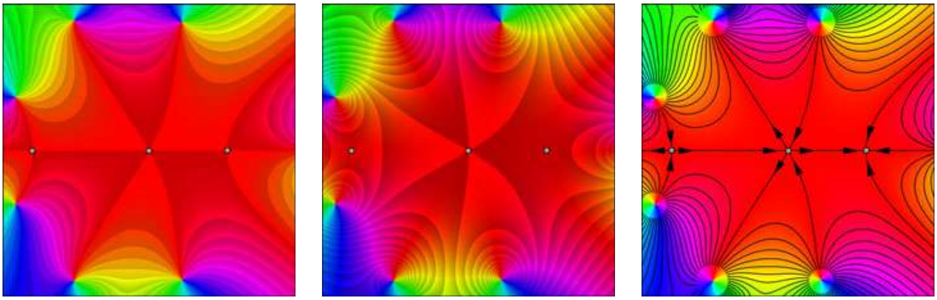


Figure 4.8: Analysis of a phase portrait with saddle points

The figure shows phase portraits of a function with three saddle points (marked by gray dots). The saddle point in the middle has order 2, the two others are of order 1. In order to localize the zero z_0 of the derivative, the color scheme in the left window has been modified by superimposing a gray component with a jump at the point $t := f(z_0)/|f(z_0)|$ of the color circle, which enhances the isochromatic lines through z_0. In the picture in the middle some contour lines are emphasized. Here the gray shading is chosen such that *in between two enhanced contour lines* greater values of $|f|$ correspond to brighter colors. *On the enhanced contour lines* the function increases towards the brighter side, and decreases towards the darker side. The right window shows the isochromatic lines through the saddle points with arrows indicating the direction of increasing $|f|$.

4.2 Complex Integration

Since a complex (analytic) function f typically does not live on an interval, but in a subset D of the complex plane, there is a variety of possible integrals: in principle every (measurable) subset G of D, and in particular "curves" ("lines" or "contours") Γ, could serve as domain of integration (with respect to some complex measure). Three integrals over a two-dimensional set G and a curve Γ which often occur in theory and applications are

$$\iint_G f(z)\,dx\,dy, \qquad \int_\Gamma f(z)\,|dz|, \qquad \int_\Gamma f(z)\,dz. \tag{4.20}$$

In the first integral f is integrated over G with respect to the *area measure*. Because this can be done separately for the real and imaginary parts of f, it easily reduces to two reals integrals. The same holds for the second integral, where integration is with respect to the *arc length* of Γ. Both these integrals are not of interest in the context of this section, which is devoted exclusively to a variant of the last integral in (4.20).

In fact we take a slightly different point of view, by considering integrals *along paths* instead of line integrals *along curves*. This concept is a little simpler (and even somewhat more flexible), and it helps us to keep things apart, like a path γ, its trace $[\gamma]$, and the generated curve Γ. And with an appropriate definition of curves as equivalence classes of paths, all assertions made for path integrals will be valid for curve (line, contour, etc.) integrals as well, because they are independent of the chosen parametrization.

So the central theme of this section is integration of analytic functions along paths in the domain set, and this leads to two intertwined topics of investigation. The first one concerns reversing the operation of differentiation, i.e., the existence of a *primitive* (or *anti-derivative*). The solution of this problem is closely related to the second topic, namely the *Cauchy integral theorem.*

The construction of a primitive involves three steps: *locally* it can be represented by function elements, which are thereafter combined by analytic continuation to a *primitive along paths.* The *global existence* of a primitive then depends on the topology of the domain: if the domain is multiply connected, there may be conflicts between primitives along different paths. This inherent problem is responsible for the notion of *multiple-valued functions*, and leads naturally to the concept of *Riemann surfaces.*

The Cauchy integral theorem is a statement about path independence of integrals involving analytic functions. We derive it in three different versions of increasing complexity. In Theorem 4.2.31 we consider paths in simply connected domains; Theorem 4.2.32 is concerned with homotopic paths in arbitrary open sets. The most general version will be shown in Theorem 4.3.6 in the next section.

Integrals of Complex Functions. We begin with the simplest situation, where a complex-valued function f is to be integrated over a finite interval $[a, b]$. Splitting f into its real and imaginary parts u and v, this case can easily be reduced to the integration of real-valued functions.

Definition 4.2.1. Let $f = u + \mathrm{i}v : [a, b] \to \mathbb{C}$ be a continuous function. Then the integral of f over $[a, b]$ is defined by

$$\int_a^b f(t)\,dt := \int_a^b u(t)\,dt + \mathrm{i} \int_a^b v(t)\,dt.$$

The continuity assumption was made only for the sake of simplicity, the definition can be extended to functions which are integrable in the Riemann or Lebesgue sense as well.

Example 4.2.1. Let $k \in \mathbb{Z}$. Then

$$\int_0^{2\pi} \mathrm{e}^{\mathrm{i}kt}\,dt = \int_0^{2\pi} \cos\,kt\,dt + \mathrm{i} \int_0^{2\pi} \sin\,kt\,dt = \left\{ \begin{array}{cl} 2\pi & \text{if } \ k = 0 \\ 0 & \text{otherwise} \end{array} \right. . \tag{4.21}$$

Integral Representation of the Taylor Coefficients. As an application we establish an integral formula for the Taylor coefficients of an analytic function.

Theorem 4.2.2. *Let f be the sum of the power series $\sum a_k\,(z - z_0)^k$ with radius of convergence R. Then, for any r with $0 < r < R$,*

$$a_k = \frac{1}{2\pi r^k} \int_0^{2\pi} f\big(z_0 + r\mathrm{e}^{\mathrm{i}t}\big)\,\mathrm{e}^{-\mathrm{i}kt}\,dt, \qquad k = 0, 1, 2, \ldots. \tag{4.22}$$

Proof. Using the representation $(z - z_0)^n = r^n \mathrm{e}^{\mathrm{i}nt}$, the uniform convergence of the power series on the circle $\{z \in \mathbb{C} : |z - z_0| = r\}$, and (4.21), we get

$$\int\limits_0^{2\pi} f\big(z_0 + r\mathrm{e}^{\mathrm{i}t}\big)\,\mathrm{e}^{-\mathrm{i}kt}\,dt = \int\limits_0^{2\pi} \sum_{n=0}^{\infty} a_n\,r^n \mathrm{e}^{\mathrm{i}(n-k)t}\,dt = \sum_{n=0}^{\infty} a_n\,r^n \int\limits_0^{2\pi} \mathrm{e}^{\mathrm{i}(n-k)t}\,dt = 2\pi\,a_k\,r^k.$$

□

For $k = 0$ we immediately obtain the following corollary, which tells us that the value of the analytic function f at the center z_0 of the disk $|z - z_0| < r$ is equal to the *integral mean* of its values on the boundary of that disk.

Corollary 4.2.3 (Mean Value Theorem)**.** *Under the assumptions of Theorem 4.2.2, the analytic function f has the* mean value property

$$f(z_0) = \frac{1}{2\pi} \int_0^{2\pi} f\big(z_0 + r\mathrm{e}^{\mathrm{i}t}\big)\, dt. \tag{4.23}$$

Estimates of Integrals. Later we shall often need estimates for complex integrals. A standard estimate for *real-valued* continuous functions f on $[a,b]$ says that

$$\left| \int_a^b f(t)\, dt \right| \le \left| \int_a^b |f(t)|\, dt \right| \le (b-a) \max_{t\in[a,b]} |f(t)|. \tag{4.24}$$

In order to extend this estimate to complex-valued functions, we write the value of the integral in its polar representation,

$$\int_a^b f(t)\, dt = r\, \mathrm{e}^{\mathrm{i}\varphi}, \qquad r := \left| \int_a^b f(t)\, dt \right|, \quad \varphi \in \mathbb{R},$$

where $\varphi \in \mathbb{R}$ can be chosen arbitrarily if $r = 0$. Then

$$\left| \int\limits_a^b f(t)\, dt \right| = r = \mathrm{Re} \int\limits_a^b \mathrm{e}^{-\mathrm{i}\varphi} f(t)\, dt = \int\limits_a^b \mathrm{Re}\, \big(\mathrm{e}^{-\mathrm{i}\varphi} f(t)\big)\, dt. \tag{4.25}$$

From $\mathrm{Re}\, z \le |z|$ and estimate (4.24), applied to the real function $|f|$, we infer that

$$\int\limits_a^b \mathrm{Re}\, \big(\mathrm{e}^{-\mathrm{i}\varphi} f(t)\big)\, dt \le \int\limits_a^b \big|\mathrm{e}^{-\mathrm{i}\varphi} f(t)\big|\, dt = \int\limits_a^b \big|f(t)\big|\, dt \le (b-a) \max_{t\in[a,b]} |f(t)|. \tag{4.26}$$

Combining (4.25) and (4.26) we finally obtain (4.24) for all continuous functions $f : [a,b] \to \mathbb{C}$.

Applying this result to the mean value formula (4.23) yields an alternative proof of the maximum principle. Moreover it is a simple exercise to derive the Cauchy estimate (Lemma 3.2.4) from Theorem 4.2.2.

Path Integrals. In the next step we extend the concept of complex integration to integrals along paths in the complex plane. In order to define such integrals we need some regularity assumptions.

Definition 4.2.4. A path $\gamma : [\alpha, \beta] \to \mathbb{C}$ is called *smooth*, if γ is continuously differentiable. A path is said to be *piecewise smooth*, if it is the sum of smooth paths, $\gamma = \gamma_1 \oplus \ldots \oplus \gamma_n$.

Depending on the context, the derivative of a smooth path $t \mapsto \gamma(t)$ with respect to the real parameter t is denoted by γ', $\dot\gamma$, or $\dfrac{d\gamma}{dt}$, respectively.

The following definition reduces complex integration along a smooth path to integration on its parameter interval.

Definition 4.2.5. Suppose that $\gamma : [\alpha, \beta] \to \mathbb{C}$ is a smooth path and let $f : [\gamma] \to \mathbb{C}$ be continuous on the trace of γ. Then the integral of f along γ is defined by

$$\int_\gamma f(z)\,dz := \int_\alpha^\beta f\big(\gamma(t)\big)\,\frac{d\gamma}{dt}(t)\,dt. \tag{4.27}$$

An integral along a piecewise smooth path $\gamma = \gamma_1 \oplus \ldots \oplus \gamma_n$ is defined as the sum of the integrals along the smooth components γ_k of γ,

$$\int_\gamma f(z)\,dz := \int_{\gamma_1} f(z)\,dz + \ldots + \int_{\gamma_n} f(z)\,dz.$$

Contour Integrals. Note that we define the integral of a function *along the path* γ and *not* along its *trace* $[\gamma]$. This avoids complications which may occur when (parts of) the trace are run through several times by the point $\gamma(t)$. If one wishes to define an integral along the oriented trace of a path (usually denoted as *contour integral* or *curve integral*) it is essential that the definition (4.27) is *invariant* with respect to reparametrization. This implies, in particular, that a linear substitution $t \mapsto at+b$ does not change the value of the integral on the right-hand side of (4.27). However, if a is negative, the lower limit of the transformed parameter interval is getting larger than its upper limit. This is the reason why the integral along the reversed path γ^- satisfies

$$\int_{\gamma^-} f(z)\,dz = -\int_\gamma f(z)\,dz.$$

The invariance of the integral with respect to smooth reparametrization of γ can be shown directly using the substitution rules. For *analytic* integrands we obtain a much more general result in Theorem 4.2.32.

Example 4.2.2. The formula for the Taylor coefficients a_k in Theorem 4.2.2 can be written as an integral along the positively oriented circle $|z - z_0| = r$. With the standard parametrization $\gamma(t) = z_0 + r\,\mathrm{e}^{\mathrm{i}t}$ we get

$$\int\limits_\gamma \frac{f(z)}{(z-z_0)^{k+1}}\,dz = \int\limits_0^{2\pi} \frac{f(z + r\,\mathrm{e}^{\mathrm{i}t})}{r^{k+1}\mathrm{e}^{\mathrm{i}(k+1)t}}\,r\,\mathrm{i}\,\mathrm{e}^{\mathrm{i}t}\,dt = \frac{\mathrm{i}}{r^k}\int\limits_0^{2\pi} f\big(z_0 + r\mathrm{e}^{\mathrm{i}t}\big)\,\mathrm{e}^{-\mathrm{i}kt}\,dt = 2\pi\mathrm{i}\,a_k. \tag{4.28}$$

Winding Numbers Revisited. It is interesting that the *winding number* of a piecewise smooth closed path (loop) can also be defined by an integral (compare Definition 2.7.20).

Lemma 4.2.6. *Let γ be a piecewise smooth closed path. Then for all $z_0 \in \mathbb{C} \setminus [\gamma]$*

$$\operatorname{wind}(\gamma, z_0) = \frac{1}{2\pi\mathrm{i}} \int_\gamma \frac{dz}{z - z_0}. \tag{4.29}$$

Proof. We assume that $\gamma : [0,1] \to \mathbb{C}$ is smooth, the general case requires only minor modifications. The function $t \mapsto \gamma(t) - z_0$ is continuous and non-vanishing on $[0,1]$. Thus, by Lemma 2.7.19, there exist continuous functions $a : [0,1] \to \mathbb{R}$ and $r : [0,1] \to \mathbb{R}_+$ such that $\gamma(t) = z_0 + r(t) \exp\big(\mathrm{i}\, a(t)\big)$ for all $t \in [0,1]$. Since γ is smooth, a and r are continuously differentiable. Computing the derivative $\dot\gamma$ of γ with respect to t we find $\dot\gamma = \big(\dot r + \mathrm{i}\, r\, \dot a\big)\exp(\mathrm{i}\, a)$. Consequently, by the fundamental theorem of real calculus,

$$\int_\gamma \frac{1}{z - z_0}\, dz = \int_0^1 \frac{\dot\gamma(t)}{\gamma(t) - z_0}\, dt = \int_0^1 \left[\frac{\dot r(t)}{r(t)} + \mathrm{i}\, \dot a(t)\right] dt = \int_0^1 \frac{d}{dt}\big(\log r + \mathrm{i}\, a\big)(t)\, dt$$
$$= \log r(1) - \log r(0) + \mathrm{i}\,\big(a(1) - a(0)\big).$$

Now the desired result follows, since $r(1) = |\gamma(1)| = |\gamma(0)| = r(0)$, and by definition of the winding number, $a(1) - a(0) = 2\pi \,\mathrm{wind}\,(\gamma, z_0)$. ☐

The Length of a Path. Though a path is a function, we associate a length with it. In fact the length of a path γ defined below coincides with the Euclidean length of its trace if no part of $[\gamma]$ is run through more than once.

Definition 4.2.7. The *length* of a smooth path $\gamma : [\alpha, \beta] \to \mathbb{C}$ is defined by

$$L(\gamma) := \int_\alpha^\beta \left|\frac{d\gamma}{dt}(t)\right| dt,$$

and the length of a piecewise smooth path is the sum of the lengths of its smooth components.

A Standard Estimate of Integrals. The following two technical results will often be needed later. The first one is an integral estimate which extends (4.24).

Lemma 4.2.8 (Standard Integral Estimate)**.** *Let* γ *be a piecewise smooth path of length* $L(\gamma)$*, and assume that* $f : [\gamma] \to \mathbb{C}$ *is continuous. If* $M(f)$ *denotes the maximum of* $|f|$ *on* $[\gamma]$*, then*

$$\left|\int_\gamma f(z)\, dz\right| \le L(\gamma) \cdot M(f).$$

Proof. Assuming in the first step that γ is smooth, we estimate the integral using (4.24),

$$\left|\int_\gamma f(z)\, dz\right| = \left|\int_\alpha^\beta f\big(\gamma(t)\big)\, \frac{d\gamma}{dt}(t)\, dt\right| \le \int_\alpha^\beta M(f) \left|\frac{d\gamma}{dt}(t)\right| dt = M(f)\, L(\gamma).$$

If $\gamma = \gamma_1 \oplus \ldots \oplus \gamma_n$ is piecewise smooth, the result follows by adding up the corresponding estimates for the integrals along the smooth paths γ_k. ☐

Lemma 4.2.9. *Let γ be a piecewise smooth path, and assume that the sequence (f_n) of continuous functions $f_n : [\gamma] \to \mathbb{C}$ converges uniformly on $[\gamma]$ to f. Then*

$$\lim_{n\to\infty} \int_\gamma f_n(z)\,dz = \int_\gamma f(z)\,dz.$$

Proof. Note that the limit function f is continuous on $[\gamma]$. The result now follows from Lemma 4.2.8 since the maximum of $f - f_n$ on γ tends to zero as $n \to \infty$. ☐

Reversing Differentiation. After these preparations we can turn our attention to the main problem of this section: is it possible to reverse the differentiation of a function? The next definition gives a more precise statement of what this means.

Definition 4.2.10. Let $f : D \to \mathbb{C}$ be a complex function on the open set D. A function $F : D \to \mathbb{C}$ is a *primitive* for f in D, if F is differentiable in D and $F' = f$ in D.

A synonymous notion for primitive is "*anti-derivative*". We point out that, in this definition as well as in the following theorem, the function f is *not assumed* to be differentiable. Later we shall see that a function f which has a primitive must *automatically* be *infinitely* differentiable.

Properties of Primitives. We put the general question of the existence of a primitive aside for a while, and look instead at what it can be used for.

Theorem 4.2.11 (Fundamental Theorem of Complex Calculus)**.** *Assume that F is a primitive of a* continuous *function $f : D \to \mathbb{C}$. Then, for any piecewise smooth path γ from a to b in D,*

$$\int_\gamma f(z)\,dz = F(b) - F(a). \tag{4.30}$$

In particular the integral vanishes if γ is closed.

Proof. We may assume that γ is smooth. Let $\gamma : [\alpha, \beta] \to D$. By definition of the integral, the chain rule, and the fundamental theorem of *real* calculus,

$$\int_\gamma f(z)\,dz = \int_\alpha^\beta f\big(\gamma(t)\big)\,\frac{d\gamma}{dt}(t)\,dt = \int_\alpha^\beta \frac{d}{dt}\,F\big(\gamma(t)\big)\,dt = F\big(\gamma(\beta)\big) - F\big(\gamma(\alpha)\big).$$

☐

Corollary 4.2.12. *If the derivative of a function f vanishes in a domain D, then f is constant in D.*

Proof. By assumption, the function f is a primitive of the zero function. Since D is a domain, any two points z_0 and z can be joined by a smooth path γ in D (see Proposition 2.7.13), and by Theorem 4.2.11

$$f(z) - f(z_0) = \int_\gamma f'(z)\,dz = 0.$$

□

Corollary 4.2.13. *If F_0 is a primitive of a function f in a domain D, then precisely the functions $F = F_0 + C$ with constant $C \in \mathbb{C}$ are primitives of f in D.*

Proof. It is clear that $F_0 + C$ is a primitive of f in D. Conversely, if F_0 and F are two such functions, then $(F - F_0)' = f - f = 0$ in D and then $F - F_0 = \text{const}$ by Corollary 4.2.12. □

Example 4.2.3. The principal branch of the complex logarithm $F(z) = \log z$ is a primitive of $f(z) = 1/z$ in $D := \{z \in \mathbb{C} : |z - 1| < 1\}$ with $F(1) = 0$ (see Example 4.1.3). Computing the integral of f along the path γ from $a = 1$ to $b = r\,\mathrm{e}^{\mathrm{i}\varphi}$ which is composed by the segment $[1, r]$ and the circular arc from r to z with center at the origin, we get from Theorem 4.2.11 that

$$F(b) = \int_\gamma f(z)\,dz = \int_1^r \frac{1}{x}\,dx + \int_0^\varphi \frac{\mathrm{i}r\,\mathrm{e}^{\mathrm{i}t}}{r\,\mathrm{e}^{\mathrm{i}t}}\,dt = \log r + \mathrm{i}\,\varphi = \log|b| + \mathrm{i}\,\mathrm{Arg}\,b.$$

Primitives of Power Series. The crucial question we would like answer is about the *existence* of a primitive. The following lemma gives a first positive result for the sum of a convergent power series

$$f(z) := \sum_{k=0}^{\infty} a_k\,(z - z_0)^k, \quad z \in D. \tag{4.31}$$

Lemma 4.2.14. *Let f be the sum of the power series (4.31) in the disk of convergence D. Then f has a primitive in D, namely*

$$F(z) = \sum_{k=0}^{\infty} \frac{a_k}{k+1}\,(z - z_0)^{k+1}, \quad z \in D.$$

Proof. By the Cauchy-Hadamard criterion (Theorem 3.2.1), the series for f and F have the same radius of convergence, and then $F' = f$ follows from Theorem 4.1.4 and (4.4). □

Example 4.2.4 (Fresnel Integrals). The power series of the entire functions given by $c(z) := \cos\left(z^2\right)$ and $s(z) := \sin\left(z^2\right)$ at $z_0 = 0$ can be derived from the series for the cosine and the sine functions by substituting z with z^2,

$$c(z) = \sum_{k=0}^{\infty}(-1)^k \frac{z^{4k}}{(2k)!}, \qquad s(z) = \sum_{k=0}^{\infty}(-1)^k \frac{z^{4k+2}}{(2k+1)!}.$$

Applying Lemma 4.2.14 we obtain power series representations of their primitives,

$$C(z) = \sum_{k=0}^{\infty}(-1)^k \frac{z^{4k+1}}{(2k)!(4k+1)}, \qquad S(z) = \sum_{k=0}^{\infty}(-1)^k \frac{z^{4k+3}}{(2k+1)!(4k+3)}.$$

The entire functions C and S are called *Fresnel integrals* and arise in the description of *diffraction phenomena* in optics. Figure 4.9 show phase portraits of C (left) and S (right) in the square $|\mathrm{Re}\, z| < 5$, $|\mathrm{Im}\, z| < 5$. It is remarkable that both functions have finite limits as $z \to \infty$ along the real or the imaginary axis, while they are rapidly growing along all other rays emerging from the origin.

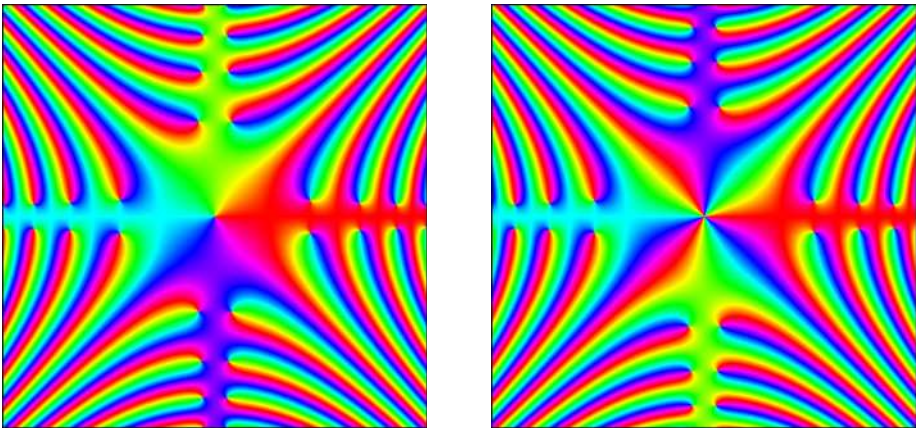


Figure 4.9: Primitives of $\cos z^2$ and $\sin z^2$ (Fresnel integrals)

Existence of Primitives. The general question about the existence of a primitive will be answered in several steps. We start with a simple situation and state a sufficient criterion. Recall that $\triangle(z_1, z_2, z_3)$ denotes the triangle with vertices z_1, z_2, z_3. An integral along the boundary $\partial\triangle(z_1, z_2, z_3)$ is, by definition, an integral along the path $[z_1, z_2] \oplus [z_2, z_3] \oplus [z_3, z_1]$.

Lemma 4.2.15. *Let $f : D \to \mathbb{C}$ be a continuous function in a disk D. If the integral of f along the boundary of any triangle $\triangle$ in D vanishes, then f has a primitive in D.*

Proof. We denote the center of D by a and define $F(z)$ in D as the integral of f along a straight line from a to z,

$$F(z) := \int_{[a,z]} f(w)\, dw.$$

Let z_0 and z be arbitrary points in D. Applying the assumption to the triangle $\triangle(z_0, a, z)$ we get

$$0 = \int\limits_{[z_0,a]} f(w)\, dw + \int\limits_{[a,z]} f(w)\, dw + \int\limits_{[z,z_0]} f(w)\, dw = -F(z_0) + F(z) + \int_{[z,z_0]} f(w)\, dw,$$

such that for $z \neq z_0$,

$$\frac{F(z)-F(z_0)}{z-z_0} - f(z_0) = \frac{1}{z-z_0} \int_{[z_0,z]} \big(f(w)-f(z_0)\big)\,dw.$$

By the standard integral estimate we obtain

$$\left|\frac{F(z)-F(z_0)}{z-z_0} - f(z_0)\right| \le \frac{|z-z_0|}{|z-z_0|} \max_{w\in[z_0,z]} |f(w)-f(z_0)|.$$

Because f is continuous in D, the right-hand side tends to zero as $z \to z_0$. Thus F is differentiable at z_0 and $F'\big(z_0\big) = f(z_0)$. □

Goursat's Lemma. The next result is a real surprise. This essential fact was discovered relatively late (by Édouard Goursat in 1884) and was not known to the founding fathers of the theory. We give a formulation which is due to Alfred Pringsheim.

Lemma 4.2.16 (Goursat)**.** *If f is differentiable in the open set D and $\triangle$ is any triangle in D, then the integral of f along the boundary of $\triangle$ vanishes.*

Proof. 1. We set $\triangle_0 := \triangle$ and denote by δ_0 the standard parametrization of its boundary (see page 57). The three segments connecting the midpoints of the sides of $\triangle_0$ divide this triangle into four sub-triangles which we denote by $\triangle_1^j$, with $j = 1, 2, 3, 4$. If δ_1^j is the standard parametrization of the boundary $\partial\triangle_1^j$, we have (see Figure 4.10)

$$\int_{\delta_0} f(z)\,dz = \int_{\delta_1^1} f(z)\,dz + \int_{\delta_1^2} f(z)\,dz + \int_{\delta_1^3} f(z)\,dz + \int_{\delta_1^4} f(z)\,dz. \tag{4.32}$$

Consequently, there must exist an index $j \in \{1, 2, 3, 4\}$ such that

$$\left|\int_{\delta^0} f(z)\,dz\right| \le 4 \left|\int_{\delta_1^j} f(z)\,dz\right|.$$

We select one such index j, define $\triangle_1 := \triangle_1^j$, and denote by δ_1 the standard parametrization of its boundary. Applying the subdivision procedure again to the triangle $\triangle_1$, we find that the integral along the boundary δ_2 of one of its four sub-triangles $\triangle_2^j$ must satisfy

$$\left|\int_{\delta_0} f(z)\,dz\right| \le 4^2 \left|\int_{\delta_2} f(z)\,dz\right|.$$

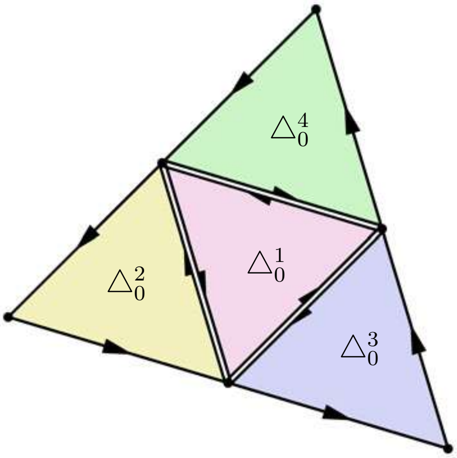


Figure 4.10: Subdivision of a triangle

Proceeding recursively in this manner, we obtain a sequence

$$\triangle_0 \supset \triangle_1 \supset \triangle_2 \supset \ldots \supset \triangle_k \supset \ldots \tag{4.33}$$

of nested triangles, such that the integrals of f along the standard parametrization δ_k of their boundaries satisfy

$$\left| \int_{\delta_0} f(z)\, dz \right| \le 4^k \left| \int_{\delta_k} f(z)\, dz \right| .$$

2. The intersection of all (compact) triangles in the family (4.33) consists of exactly one point z_0 in $\triangle$. Because $\triangle \subset D$, the function f is differentiable at z_0, that is

$$f(z) = f(z_0) + f'(z_0)(z - z_0) + r(z)$$

with $r(z) = o(z - z_0)$. Consequently, for any $\varepsilon > 0$ there exists a $\delta > 0$ such that $|r(z)| < \varepsilon \cdot |z - z_0|$ whenever $|z - z_0| < \delta$.

Let d denote the circumference of the triangle $\triangle$. Then all points z in $\triangle_k$ satisfy $|z - z_0| \le d/2^k$, so that for all sufficiently large k

$$|r(z)| < \varepsilon\, \frac{d}{2^k}, \qquad \text{if } z \in \triangle_k. \tag{4.34}$$

3. Since the linear function $z \mapsto f(z_0) + f'(z_0)(z - z_0)$ has a primitive, the integral over that part of f along the closed path δ_k vanishes. Hence, by (4.34) and the standard integral estimate of Lemma 4.2.8,

$$\left| \int_{\delta_0} f(z)\, dz \right| \le 4^k \left| \int_{\delta_k} r(z)\, dz \right| \le 4^k\, L\big(\delta_k\big)\, \varepsilon \frac{d}{2^k} = 4^k\, \varepsilon\, \frac{d}{2^k}\, \frac{d}{2^k} = \varepsilon\, d^2, \tag{4.35}$$

for all sufficiently large k. Now ε can be chosen arbitrarily small, which means that the left-hand side of (4.35) must be zero. □

Combining Lemma 4.2.16 with Lemma 4.2.15 we obtain the following important corollary.

Corollary 4.2.17. *If f is differentiable in a disk D, then it has a primitive in D.*

Differentiable and Analytic Functions. After all these preparations we can close the gap between the classes of differentiable and analytic functions. So far we know from Theorem 4.1.4 that any analytic function is differentiable. The converse is absolutely counterintuitive and has many strange consequences. Why, for example, should a function which is differentiable in an open set *automatically* have derivatives of *arbitrary order*? It belongs to the deep miracles of complex function theory that this is really so.

The proof of this fact is a bit tricky. A direct attempt would perhaps aim at proving that a holomorphic function f is infinitely differentiable, and then trying to show that it can be represented locally as the sum of a power series. Instead of doing so we go one step backward, and show that the *primitive* F of a holomorphic function f is analytic. This follows from an appropriate integral representation of F. Once we know that F is analytic, we can conclude that all derivatives of F, and in particular its first derivative f, are analytic too.

Lemma 4.2.18. *Assume that $f : D \to \mathbb{C}$ is differentiable in a disk D with center a and radius R. Fix r with $0 < r < R$ and let $\gamma(t) = a + r\,\mathrm{e}^{\mathrm{i}t}$ for $t \in [0, 2\pi]$. If F is a primitive of f, then for all points z_0 with $|z_0 - a| < r$,*

$$F(z_0) = \frac{1}{2\pi\mathrm{i}} \int_\gamma \frac{F(z)}{z - z_0}\, dz. \tag{4.36}$$

Proof. 1. We begin with an auxiliary result. Let $z_0 \in D$ be fixed. For h in the disk $D_0 := \{z \in \mathbb{C} : z + z_0 \in D\}$ we define

$$\varphi(h) := F\big(z_0 + h\big) - F\big(z_0\big) - f\big(z_0\big)\, h - \frac{1}{2}\, f'\big(z_0\big)\, h^2.$$

The function $h \mapsto \varphi(h)$ is differentiable in D_0 and its derivative with respect to h can be computed by the chain rule,

$$\varphi'(h) = F'\big(z_0 + h\big) - f\big(z_0\big) - f'\big(z_0\big)\, h = f\big(z_0 + h\big) - f\big(z_0\big) - f'\big(z_0\big)\, h.$$

Since f is differentiable at z_0, the right-hand side is of order $o(h)$ as $h \to 0$, that is, for any $\varepsilon > 0$ there exists a $\delta > 0$ such that $|\varphi'(h)| \le \varepsilon |h|$ whenever $|h| < \delta$.

The function φ' is continuous, whence the mapping $[0, 1] \to \mathbb{C},\ t \mapsto \varphi(th)$ is continuously differentiable (with respect to the real variable t), so that $\varphi(h)$ can be represented by the fundamental theorem of calculus,

$$\varphi(h) = \int_0^1 \frac{d}{dt}\big(\varphi(th)\big)\, dt = \int_0^1 \varphi'(th)\, h\, dt.$$

Using the standard estimate for integrals in combination with $|\varphi'(h)| \le \varepsilon |h|$, we conclude that $|\varphi(h)| \le h^2\, \varepsilon$ for all h with $|h| < \delta$. Since ε can be chosen arbitrarily small,

$$\lim_{h \to 0}\ \varphi(h)/h^2 = 0. \tag{4.37}$$

2. The function G defined by

$$G(z) := \begin{cases} \dfrac{F(z) - F(z_0)}{z - z_0} & \text{if} \quad z \in D \setminus \{z_0\} \\ f(z_0) & \text{if} \quad z = z_0 \end{cases}$$

is differentiable in $z \in D \setminus \{z_0\}$. In order to prove that G is also differentiable at z_0, we consider its difference quotient at z_0. By definition of G and φ,

$$\frac{G(z) - G(z_0)}{z - z_0} = \frac{1}{(z - z_0)^2}\,\big(F(z) - F(z_0) - (z - z_0)\, f(z_0)\big) = \frac{1}{2}\, f'(z_0) + \frac{\varphi(z - z_0)}{(z - z_0)^2},$$

and using (4.37) we find $G'(z_0) = (1/2)\, f'(z_0)$.

3. Because G is differentiable in the disk D, we can apply Corollary 4.2.17, which tells us that the integral of G along the closed path γ vanishes. Hence

$$0 = \int_\gamma G(z)\, dz = \int_\gamma \frac{F(z) - F(z_0)}{z - z_0}\, dz = \int_\gamma \frac{F(z)}{z - z_0}\, dz - F(z_0) \int_\gamma \frac{dz}{z - z_0}.$$

By Lemma 4.2.6, the last integral is equal to $2\pi \mathrm{i} \cdot \operatorname{wind}(\gamma, z_0) = 2\pi \mathrm{i}$. □

Cauchy Integrals. The expression on the right-hand side of formula (4.36) is said to be a *Cauchy integral*. We shall study integrals of this type in a more general setting in the next section. In view of what we will explore in Section 5.5 we consider a more general situation than in (4.36), where the function F in the integrand is replaced by an arbitrary continuous complex function φ. This setting is described in the following definition.

Definition 4.2.19. Let γ be a piecewise smooth path in $\mathbb{C}$ and assume that the function $\varphi : [\gamma] \to \mathbb{C}$ is continuous. The function $f : \mathbb{C} \setminus [\gamma] \to \mathbb{C}$ defined by

$$f(z) := \frac{1}{2\pi \mathrm{i}} \int_\gamma \frac{\varphi(w)}{w - z}\, dw, \qquad z \in \mathbb{C} \setminus [\gamma] \tag{4.38}$$

(and the integral on the right-hand side of (4.38)) is said to be the *Cauchy integral with density* φ along γ.

Here is the crucial result that will bridge the gap between differentiable and analytic functions.

Theorem 4.2.20. *Let γ be a piecewise smooth path in $\mathbb{C}$ and assume that $\varphi : [\gamma] \to \mathbb{C}$ is continuous. Then the function f defined by the Cauchy integral (4.38) is analytic on $D := \mathbb{C} \setminus [\gamma]$ and tends to zero as $z \to \infty$. For any disk $D_0 \subset D$ with center z_0 the Taylor series*

$$f(z) = \sum_{k=0}^{\infty} a_k\, (z - z_0)^k$$

of f at z_0 converges in D_0 and its coefficients satisfy

$$a_k = \frac{1}{2\pi \mathrm{i}} \int_\gamma \frac{\varphi(z)}{(z - z_0)^{k+1}}\, dz. \tag{4.39}$$

Proof. Fix $z \in D_0$. Invoking a compactness argument, we know of the existence of a constant $q < 1$ such that $|z - z_0|/|w - z_0| \le q < 1$ for all $w \in [\gamma]$. Consequently,

$$\frac{\varphi(w)}{w - z} = \frac{\varphi(w)}{(w - z_0) - (z - z_0)} = \sum_{k=0}^{\infty} \frac{\varphi(w)}{w - z_0} \left(\frac{z - z_0}{w - z_0}\right)^k.$$

Since the continuous function $w \mapsto \varphi(w)/(w-z_0)$ is bounded on the compact set $[\gamma]$, the series converges uniformly with respect to $w \in [\gamma]$. Interchanging the order of summation and integration, we obtain

$$2\pi\mathrm{i}\, f(z) = \int_\gamma \frac{\varphi(w)}{w-z}\, dw = \sum_{k=0}^{\infty} \left(\int_\gamma \frac{\varphi(w)}{(w-z_0)^{k+1}}\, dw \right) (z-z_0)^k$$

for all $z \in D_0$, which proves the claim. Finally, the standard integral estimate yields that $f(z) \to 0$ for $z \to \infty$. ☐

Differentiable Functions are Analytic. Now the hard work is done and we can harvest the fruits.

Theorem 4.2.21. *Any function $f : D \to \mathbb{C}$ which is differentiable in an open subset D of the complex plane is analytic in D.*

Proof. Since both properties are local, we can assume that D is a disk. Corollary 4.2.17 guarantees that f has a primitive F in D. By Lemma 4.2.18 the function F can be represented as a Cauchy integral, so that F is analytic (Theorem 4.2.20), which implies that $f = F'$ is analytic (Theorem 4.1.7). ☐

From Theorems 4.1.4 and 4.2.21, it follows that it makes no difference whether we speak of "differentiable" or "analytic" functions, at least if we restrict ourselves to functions $f : D \subset \mathbb{C} \to \mathbb{C}$. Moreover, "holomorphic" is just another synonym for functions in this class (see Definition 3.5.4).

In order not to confuse the reader with too many names we shall preferably speak of analytic functions, mainly because this concept has been introduced first and several results have been explicitly stated for this class. Also phrases like "analytic continuation of a holomorphic function" would sound a bit odd.

A Criterion of Analyticity. Before we continue our investigation of the existence of a primitive we obtain a number of additional results. The following criterion for analyticity is the converse of Goursat's lemma:

Theorem 4.2.22 (Morera)**.** *Let $f : D \to \mathbb{C}$ be a continuous function on an open set $D \subset \mathbb{C}$. If the integral of f along the boundary of any triangle $\triangle$ in D vanishes, then f is analytic in D.*

Proof. We may assume that D is a disk. Then, by Lemma 4.2.15, the function f has a primitive F in D. By definition of the primitive, this function is differentiable and hence analytic, which implies that $f = F'$ is analytic. ☐

Another surprising result, which has no counterpart in real analysis, shows that the analyticity of continuous functions is not disturbed by "thin" exceptional sets. We state a version which can easily be derived from Morera's theorem. In the following corollary, the notion of line segments also includes points.

Corollary 4.2.23. *Let D be an open subset of the plane and let S be a finite union of line segments in D. If the function f is continuous in D and analytic in $D \setminus S$, then f is analytic in D.*

Disk of Convergence Revisited. Another remarkable fact is that the Taylor series at z_0 of an analytic function f always converges (at least) in the largest disk with center z_0 which fits into the domain set of f.

Theorem 4.2.24. *Let f be analytic in a disk D with center z_0. Then the Taylor series of f at z_0 converges in the entire disk D.*

Proof. We denote the radius of the disk D by R and fix an r with $0 < r < R$. Then the path $\gamma_r(t) = z_0 + r\mathrm{e}^{\mathrm{i}t}$ lies in D and $|f|$ is bounded on γ_r by some constant M. Applying the the standard integral estimate to the coefficient formula (4.39), we obtain the estimates

$$|a_k| \le M/r^k, \qquad k = 0, 1, 2, \ldots . \tag{4.40}$$

By Lemma 3.2.2 the Taylor series converges in the disk $|z - z_0| < r$, and because r can be chosen arbitrarily close to R it must converge everywhere in D. ☐

We point out that the above result is not a tautology: being analytic in a disk means that f can *locally* be represented as the sum of a power series, but there is no *obvious* reason why one such series should converge *globally* in the entire disk.

Bounded Entire Functions. The coefficient estimate (4.40) has another astounding consequence.

Theorem 4.2.25 (Liouville)**.** *Any bounded entire function is constant.*

Proof. If f is analytic in the complex plane and $|f| \le M$, the estimate (4.40) for the Taylor coefficients a_k applies for every $r > 0$, so that $a_k = 0$ for $k = 1, 2, \ldots$. ☐

Global Existence of Primitives. By establishing Theorem 4.2.21 we have achieved one of our main goals. However, the existence of a primitive on general domains is still open, Corollary 4.2.17 settles it only for differentiable functions in a disk. Indeed problems may arise when we try to globalize this result.

Example 4.2.5. The function $f(z) = 1/z$ is analytic in $D := \mathbb{C} \setminus \{0\}$. If f had a primitive in D, then the integral of f over any path γ in D would vanish by Theorem 4.2.11. However, for the closed path γ defined on $[0, 2\pi]$ by $\gamma(t) := \mathrm{e}^{\mathrm{i}t}$ we have (see Example 4.2.1)

$$\int_\gamma \frac{dz}{z} = 2\pi\mathrm{i} \ne 0.$$

The function $1/z$ plays an exceptional role among all power functions $f(z) = z^n$ with integer exponent n. Indeed, $F(z) = z^{n+1}/(n+1)$ is a primitive of f for all $n \ne -1$.

Primitives Along Paths. Since, at least in some situations, there is no hope that a primitive exists globally on the entire domain of an analytic function, we try to find an appropriate substitute. The following construction goes by the name "primitive along a path": Take a path γ, cover it by a chain of disks and glue together the patches of primitives of f in the covering disks. In other words, we build the *analytic continuation* of a primitive along γ.

Definition 4.2.26. A *primitive of f along* a path γ is a chain of function elements $\big(F_0, D_0\big) \infty \big(F_1, D_1\big) \infty \ldots \infty \big(F_n, D_n\big)$ which covers the path γ and satisfies the condition $F_k' = f$ on D_k for $k = 0, \ldots, n$.

The next theorem claims that this procedure always works. Note that we do not require any regularity of the underlying paths.

Theorem 4.2.27 (Analytic Continuation of Primitive)**.** *Let $f : D \subset \mathbb{C} \to \mathbb{C}$ be analytic in the open set D and let (F_0, D_0) be a primitive of f in the disk $D_0 \subset D$. Then (F_0, D_0) has an unrestricted analytic continuation in D, and all resulting function elements are primitives of f on their domain sets.*

Proof. According to the definition of an unrestricted analytic continuation (see page 125) we must show that the function element (F_0, D_0) has an analytic continuation F_γ along any path γ in D which starts at the center z_0 of D_0. By the path covering lemma (Lemma 2.7.8) there exists a chain of disks $(D_0, D_1, \ldots, D_n)$ which covers γ in the sense of Definition 2.7.7. Using Corollary 4.2.17 we find function elements $\big(F_0, D_0\big), \ldots, \big(F_n, D_n\big)$ such that F_k is a primitive of f in D_k for $k = 0, 1, \ldots, n$.

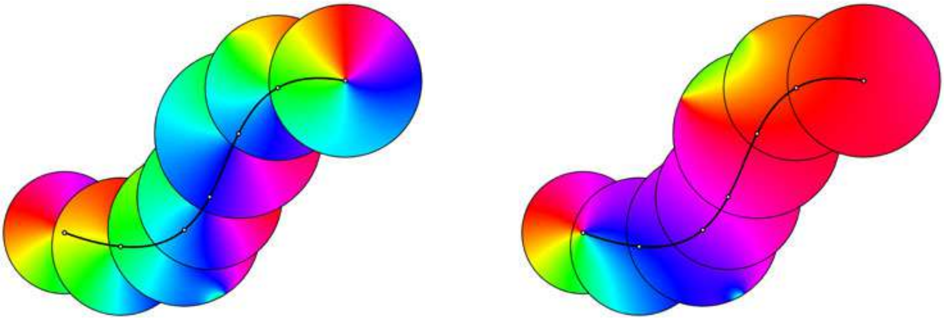


Figure 4.11: Normalized and adjusted function elements of primitives along a path

Corollary 4.2.13 tells us that $F_k - F_{k-1}$ is constant on $D_{k-1} \cap D_k$, so that there exist constants $C_0 := 0$ and $C_1, \ldots, C_n$ such that $F_{k-1} + C_{k-1} = F_k + C_k$ on $D_{k-1} \cap D_k$ for $k = 1, \ldots, n$. Consequently, the chain of function elements

$$\big(F_0, D_0\big) \infty \big(F_1 + C_1, D_1\big) \infty \ldots \infty \big(F_n + C_n, D_n\big)$$

is an analytic continuation of (F_0, D_0) along γ composed of primitives of f in the covering disks D_k. $\square$

Figure 4.11 shows a path γ covered by a family of disks. Every disk D_k carries the phase portrait of a primitive F_k of some function f. In the picture on the left the primitives are normalized such that F_k vanishes at the center of D_k. In the right we start with the same function F_0 near the initial point of the path, but the primitives in the other disks are adjusted such they fit together to form a primitive of f along γ.

Existence of Global Primitives. In combination with the monodromy principle, the above theorem provides an affirmative answer to the question about the existence of a primitive (as a single-valued function) for a special class of domains.

Theorem 4.2.28. *If f is analytic in the simply connected domain D, then f has a primitive on D.*

Proof. Fix a disk $D_0 \subset D$ with center z_0. Then f has a primitive F_0 in D_0. By Theorem 4.2.27 and assertion (ii) of the monodromy principle (Theorem 3.6.10), the analytic continuation of the function element (F_0, D_0) along all paths γ with (fixed) initial point z_0 and (variable) endpoint z in D is a function F of z. Since all function elements involved in the analytic continuation are primitives of f (in their respective domains) F is a primitive of f in D. ☐

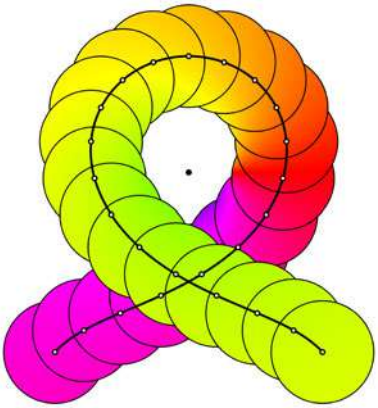


Figure 4.12: Continuation of a primitive

In general the situation is more complicated, and even the idea of using the function elements (F_k, D_k) covering γ to construct a primitive of f as function of z in a neighborhood of the *trace of a path* γ may fail if γ is not simple. This is demonstrated in Figure 4.12 for the function $f(z) = 1/z$. Every function F_k is a primitive of $1/z$ in the corresponding disk D_k and, starting at the initial point of γ, neighboring function elements are adjusted such that they coincide in the overlapping regions of their domains. After travelling around the origin along γ, a conflict arises when we visit the same region again.

In those cases where a primitive in a neighborhood of $[\gamma]$ does not exist, we can get a substitute by *pulling back* the functions F_k from the disks D_k to the parameter interval $[\alpha, \beta]$ of γ,

$$F_\gamma(t) := F_k\big(\gamma(t)\big), \qquad t \in [t_{k-1}, t_k], \quad k = 1, \dots, n \tag{4.41}$$

(for a description of the intervals $[t_{k-1}, t_k]$ see Definition 2.7.7). By Lemma 3.6.5, this definition does not depend on the choice of the covering chain, and we shall refer to the function $F_\gamma : [\alpha, \beta] \to \mathbb{C}$ also as a *primitive of* f *along* γ. The next result shows that this definition is perfectly suited to the calculation of integrals.

Theorem 4.2.29 (Extended Fundamental Theorem). *Let* $\gamma : [\alpha, \beta] \to D$ *be a piecewise smooth path. If* f *is holomorphic in* D *and* F_γ *is a primitive of* f *along* γ *then*

$$\int_\gamma f(z)\, dz = F_\gamma(\beta) - F_\gamma(\alpha). \tag{4.42}$$

Proof. We choose a chain of function elements (F_k, D_k) which covers γ. Then the restriction γ_k of γ to the subdivision interval $[t_{k-1}, t_k]$ maps this interval into the disk D_k, and γ is decomposed as $\gamma = \gamma_1 \oplus \gamma_2 \oplus \ldots \oplus \gamma_n$. The result then follows on applying Theorem 4.2.11 to the integrals along γ_k and adding up. ☐

Integrals Along Arbitrary Paths. Since the primitive of an analytic function f along γ exists for *all* paths in the domain of f, we can use the right-hand side of (4.42) as a *definition* of the integral on the left-hand side, even when γ is *not smooth.*

Definition 4.2.30. Let f be analytic in the open set D and let $\gamma : [\alpha, \beta] \to D$ be a path in D. Then the integral of f along γ is defined by (4.42), where F_γ is a primitive of f along γ.

Constructions Involving Families of Paths. Finally we investigate what happens when we try to construct a primitive of f as follows: fix a point $z_0 \in D$, select one path $\gamma_z : [0, 1] \to D$ from z_0 to z for every z in D (or in a subset thereof), and denote the family of paths γ_z with $z \in D$ by G. Then define $F_G(z)$ as the value of the primitive of f along the path γ_z at its terminal point z,

$$F_G(z) := F_{\gamma_z}(1).$$

We call F_G a *primitive* of f *with respect to the family* G. If f is analytic in the *simply connected* domain D then F_G does not depend on the choice of the paths γ_z (see Theorem 4.2.28), but in general the properties of F_G are influenced by the selection of γ_z as the following experiments demonstrate.

Example 4.2.6. The function $f(z) = 1/z$ from Example 4.2.5 is analytic in the punctured plane $D := \mathbb{C} \setminus \{0\}$ and thus it has a primitive along any path γ in D.

Figure 4.13 (left) shows a square Q with a phase portrait of a primitive F_G of f with respect to a family G of paths γ_z. Every path γ_z is composed of two segments as follows: the first segment starts at the lower left corner z_0 of the square Q and runs parallel to the real axis. The second segment of γ_z is parallel to the imaginary axis and connects the endpoint of the first segment with the (variable) point z. The paths γ_z which meet the pole of f at $z = 0$ are excluded from the family G.

The window on the right in Figure 4.13 shows the primitive of f with respect to another family of paths which also start at z_0, but run *first* in the direction of the imaginary axis, and *thereafter* parallel to the real axis.

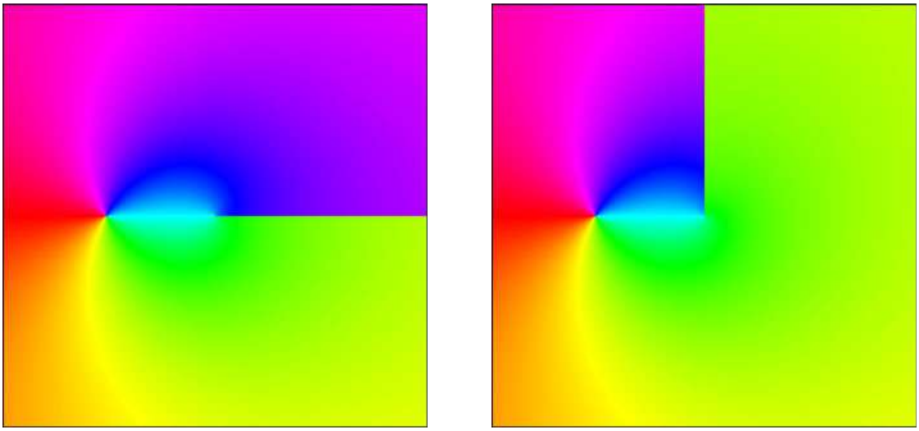


Figure 4.13: Primitives of $f(z) = 1/z$ with respect to two families of paths

In both pictures it is clearly visible that something happens “behind the poles”. Here the primitives of f along different paths are incompatible, primitives along γ_z running on the left side of the poles do not fit with those running on their right sides.

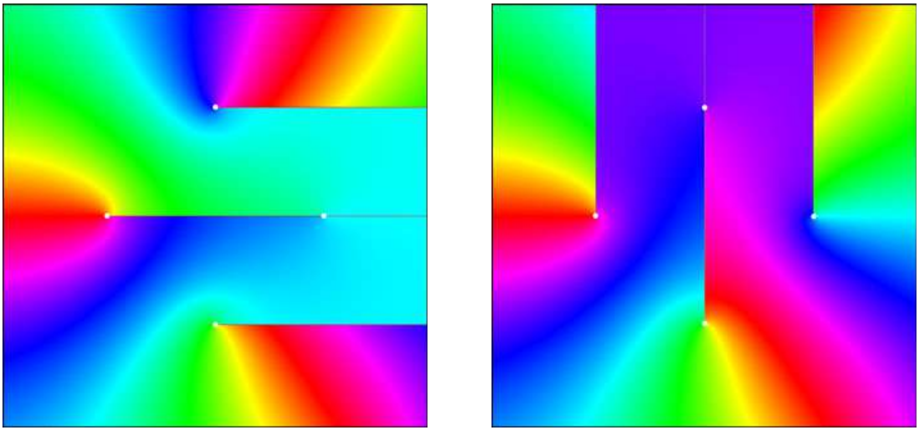


Figure 4.14: Primitives of $f(z) = 1/(z^4 - 1)$ with respect to two families of paths

Figure 4.14 shows the corresponding results for the function $f(z) = 1/(z^4 - 1)$, which is analytic in $\mathbb{C} \setminus \{1, \mathrm{i}, -1, -\mathrm{i}\}$.

Path Independence of Integrals. Let us now return to the question in which way the integral of an analytic function depends on the choice of the path of integration. For simply connected domains the answer follows immediately from Theorem 4.2.11 and Theorem 4.2.28.

Theorem 4.2.31 (Cauchy Integral Theorem I). *Let f be analytic in the simply connected domain D. If γ is a closed path in D, then*

$$\int_\gamma f(z)\,dz = 0.$$

If γ_0 and γ_1 are two paths in D from a to b, then

$$\int_{\gamma_0} f(z)\,dz = \int_{\gamma_1} f(z)\,dz.$$

The following result is more general. Here we do not require that D is simply connected. Saying that two paths γ_0 and γ_1 are homotopic, we mean that either these paths are homotopic with fixed endpoints, or that both paths are closed and freely homotopic.

Theorem 4.2.32 (Cauchy Integral Theorem II). *Let f be analytic in the open set $D \subset \mathbb{C}$. If the closed path γ is null-homotopic in D, then*

$$\int_\gamma f(z)\,dz = 0.$$

If the paths γ_0 and γ_1 are homotopic in D, then

$$\int_{\gamma_0} f(z)\,dz = \int_{\gamma_1} f(z)\,dz.$$

Proof. 1. Any primitive F_γ of f along the path $\gamma : [\alpha, \beta] \to D$ is obtained via an analytic continuation of a function element (F_0, D_0) centered at the initial point a of γ. If γ is null-homotopic in D, we infer from assertion (i) of the monodromy principle (Theorem 3.6.10) that $F_\gamma(\alpha) = F_\gamma(\beta)$, and the first assertion follows from (4.42).

2. If γ_0 and γ_1 are homotopic with fixed endpoints in D, then $\gamma := \gamma_1 \ominus \gamma_0$ is closed and null-homotopic in D (see Lemma 2.7.18). By the first step of the proof

$$\int_{\gamma_1} f(z)\,dz - \int_{\gamma_0} f(z)\,dz = \int_\gamma f(z)\,dz = 0.$$

3. Assume that γ_0 and γ_1 are closed and freely homotopic in D. We may suppose that both paths are defined on the parameter interval $[0,1]$. Then there exists a homotopy $h : [0,1] \times [0,1] \to D$ such that the family of paths γ_s, defined by $\gamma_s(t) := h(s,t)$ connects γ_0 with γ_1. For $s \in [0,1]$ we define

$$\gamma_s^+(t) := h(st, 0) = h(st, 1), \qquad \gamma_s^- := -\gamma_s^+.$$

Then $\gamma_s^* := \gamma_s^+ \oplus \gamma_s \oplus \gamma_s^-$ are closed paths in D with fixed endpoints

$$\gamma_s^*(0) = \gamma_s^*(1) = \gamma_0(0) = \gamma_0(1),$$

and the family γ_s^* is a homotopy with fixed endpoints from $\gamma_0^* = \gamma_0$ to the path $\gamma_1^* = \gamma_1^+ \oplus \gamma_1 \oplus \gamma_1^-$. Applying the result of the first step, and taking into account that the integrals along γ_1^+ and γ_1^- cancel each other, we obtain

$$\int_{\gamma_0} f\,dz = \int_{\gamma_0^*} f\,dz = \int_{\gamma_1^*} f\,dz = \int_{\gamma_1^+} f\,dz + \int_{\gamma_1} f\,dz + \int_{\gamma_1^-} f\,dz = \int_{\gamma_1} f\,dz.$$

□

The following corollary of Theorem 4.2.32 simplifies the investigation of integrals, by allowing us to replace complicated paths by very simple ones.

Theorem 4.2.33. *Let γ be a path in the open set $D \subset \mathbb{C}$. Then there exist a smooth path $\widetilde{\gamma} \subset D$ and an axis parallel polygonal path $\widehat{\gamma} \subset D$ such that for all analytic functions $f : D \to \mathbb{C}$*

$$\int_{\gamma} f(z)\,dz = \int_{\widetilde{\gamma}} f(z)\,dz = \int_{\widehat{\gamma}} f(z)\,dz.$$

Proof. The result is a direct consequence of Lemma 2.7.12 and Theorem 4.2.32. □

4.3 Cauchy Integral Formula

A real miracle happens when we apply Theorem 4.2.32 to the formula for the Taylor coefficients of analytic functions. Not only can the value of an analytic function f at a point be computed by an integral along an arbitrary curve which winds itself around that point, but we can also compute all derivatives of f by integration. This fundamental result is called the *Cauchy integral formula*. It is of crucial importance for understanding the essence of analytic functions and has countless applications.

Theorem 4.3.1 (Cauchy Integral Formula)**.** *Assume that f is* analytic *in the* simply connected *domain D and let $z_0 \in D$. If γ is a closed path in $D \setminus \{z_0\}$ with winding number one about z_0, then*

$$f(z_0) = \frac{1}{2\pi \mathrm{i}} \int_{\gamma} \frac{f(z)}{z - z_0}\,dz. \tag{4.43}$$

Moreover, for $k = 1, 2, \ldots$, the kth derivative of f at z_0 is given by the integral formula

$$f^{(k)}(z_0) = \frac{k!}{2\pi \mathrm{i}} \int_{\gamma} \frac{f(z)}{(z - z_0)^{k+1}}\,dz. \tag{4.44}$$

Proof. The Taylor coefficients a_k of f at z_0 are given by formula (4.28),

$$a_k = \frac{1}{2\pi \mathrm{i}} \int_{\gamma_r} \frac{f(z)}{(z - z_0)^{k+1}}\,dz, \tag{4.45}$$

where integration is along a positively oriented circle γ_r in the disk of convergence of that series.

Since the integrand is analytic in $D \setminus \{z_0\}$, and the closed path $\gamma \subset D \setminus \{z_0\}$ with winding number one about z_0 is freely homotopic to γ_r in the punctured domain $D \setminus \{z_0\}$ (here we need that D is simply connected, see Lemma 2.7.22), replacing the circle γ_r by γ does not change the value of the integral. To complete the proof we refer to Theorem 4.1.7, which tells us that the coefficients a_k are related to the derivatives of f by $a_k = f^{(k)}(z_0)/k!$. □

Integrals Along Chains and Cycles. Theorem 4.3.1 is by far not the end of the story for the Cauchy integral formula – it can be generalized in several directions. Here we shall state and prove one such result which admits *arbitrary domains* D and allows “paths” of integration that are composed of several pieces. We begin with some definitions.

Definition 4.3.2. A *chain* Γ (in D) is a finite collection of paths $\gamma_1, \ldots, \gamma_n$ (in D). If all paths $\gamma_1, \ldots, \gamma_n$ are closed, we call Γ a *cycle*.

Speaking of a “collection” instead of a set, we admit that some of the paths γ_k can occur several times. We say that the chain Γ is constituted by its *components* $\gamma_1, \ldots, \gamma_n$ and write

$$\Gamma = \gamma_1 + \ldots + \gamma_n. \tag{4.46}$$

The *trace* $[\Gamma]$ of a chain is the union of the traces of its components γ_k, and the *length* of Γ is the sum of the lengths of the γ_k (provided that all γ_k are piecewise smooth). The additive notation (4.46) of chains is further motivated by the definition of winding numbers

$$\operatorname{wind}(\Gamma, z) := \operatorname{wind}(\gamma_1, z) + \ldots + \operatorname{wind}(\gamma_n, z), \qquad z \notin [\Gamma],$$

and integrals,

$$\int_\Gamma f(z)\,dz := \int_{\gamma_1} f(z)\,dz + \ldots + \int_{\gamma_n} f(z)\,dz.$$

If γ is a closed path (or a cycle) in a simply connected domain D, then the winding number of γ about all points in the “exterior” $E := \mathbb{C} \setminus D$ of D is zero. This need not be so if D is multiply connected, but the subset of all cycles which have this property plays a distinguished role.

Definition 4.3.3. A cycle Γ in an open set D is said to be *null-homologous* (or *homologous to zero*) in D, if its winding number about any point in $\mathbb{C} \setminus D$ is zero,

$$\operatorname{wind}(\Gamma, z) = 0, \qquad z \in \mathbb{C} \setminus D.$$

It is clear that a cycle which is composed of *null-homotopic* paths is null-homologous, but even for a single path the converse need not be true. Two examples of null-homologous cycles in $\mathbb{C} \setminus \{-1, 1\}$ are depicted in Figure 4.15. The path on the right is null-homologous but not null-homotopic.

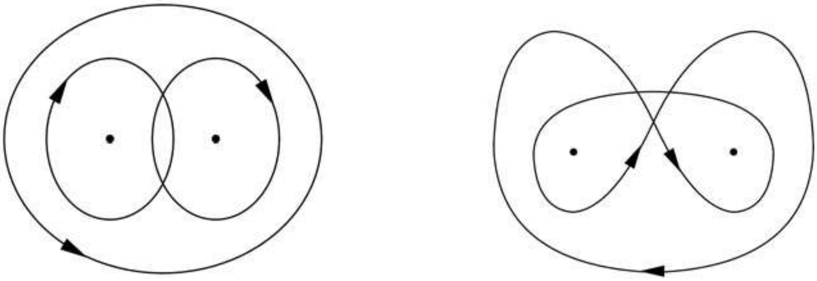


Figure 4.15: A null-homologous cycle and a null-homologous path

Cauchy Integral Formula Revisited. After these preparations we can now prove a ‘global’ version of the Cauchy integral formula.

Theorem 4.3.4 (Cauchy Integral Formula II)**.** *Let Γ be a null-homologous cycle in the open set $D \subset \mathbb{C}$. If f is analytic in D, then*

$$\frac{1}{2\pi \mathrm{i}} \int_\Gamma \frac{f(z)}{z-z_0}\, dz = \operatorname{wind}(\Gamma, z_0) \cdot f(z_0), \qquad z_0 \in D \setminus [\Gamma]. \tag{4.47}$$

Proof. 1. Let $z_0 \in D \setminus [\Gamma]$ be fixed. Since $[\Gamma]$ is compact, its distance to $\{z_0\} \cup (\mathbb{C} \setminus D)$ is positive. Using Lemma 2.7.12 and Lemma 2.7.21 on the stability of winding numbers with respect to perturbations, we find a cycle $\widehat{\Gamma}$, composed of polygonal paths $\widehat{\gamma}_k$ which are homotopic to the corresponding components γ_k of Γ in $D \setminus \{z_0\}$, and satisfies

$$\operatorname{wind}\big(\widehat{\Gamma}, z\big) = \operatorname{wind}\big(\Gamma, z\big), \qquad z \in \{z_0\} \cup \big(\mathbb{C} \setminus D\big).$$

In particular $\widehat{\Gamma}$ is null-homologous in D and both sides of (4.47) remain unchanged when Γ is replaced by $\widehat{\Gamma}$. Consequently we may assume that Γ is already composed of polygonal paths.

2. In order to verify (4.47) we consider the function g_0, defined in the domain $D_0 := D \setminus [\Gamma]$ by

$$g_0(z) := \frac{1}{2\pi \mathrm{i}} \int_\Gamma \frac{f(w)}{w-z}\, dw - \operatorname{wind}(\Gamma, z) \cdot f(z), \qquad z \in D_0. \tag{4.48}$$

Our goal is to prove that g_0 can be analytically extended to an entire function g tending to zero at infinity, so that Liouville’s theorem then implies $g = 0$ on $\mathbb{C}$. This will be realized in several steps.

3. The first summand on the right-hand side of (4.48) is the sum of Cauchy integrals along paths which are contained in $[\Gamma]$, so it can be extended to an analytic function g_1 in $\mathbb{C} \setminus [\Gamma]$,

$$g_1(z) := \frac{1}{2\pi \mathrm{i}} \int_\Gamma \frac{f(w)}{w-z}\, dw, \qquad z \in G := \mathbb{C} \setminus [\Gamma]. \tag{4.49}$$

To convince ourselves that the second summand $g_2(z)$ defines an analytic function in D_0, we remark that f is analytic in D_0 and $\operatorname{wind}(\Gamma, z)$ is constant on the components of D_0. In particular we have $g_2(z) = 0$ whenever $z \in D_0$ and $\operatorname{wind}(\Gamma, z) = 0$. The set E of all points $z \in \mathbb{C}$ for which $\operatorname{wind}(\Gamma, z) = 0$ is open and, since Γ is null-homologous in D, the complement of D is contained in E. Since g_2 is zero on the intersection of D_0 and E, it can be extended to an analytic function on G by setting $g_2(z) := 0$ if $z \in \mathbb{C} \setminus D$, so that

$$g_2(z) = \begin{cases} -\operatorname{wind}(\Gamma, z) \cdot f(z) & \text{if} \quad z \in D_0 \\ 0 & \text{if} \quad z \in E \end{cases} . \tag{4.50}$$

Finally, we define the analytic function g on G by $g := g_1 + g_2$. Note that g_0 is the restriction of g to D_0.

4. In order to prove that g has an analytic extension from $G = \mathbb{C} \setminus [\Gamma]$ to the entire plane $\mathbb{C}$, we use Corollary 4.2.23 to Morera's theorem. Since Γ is supposed to be polygonal, its trace $[\Gamma]$ is the union of a finite number of line segments. So, by virtue of Corollary 4.2.23, we only need to show that g can be extended to a *continuous* function on $\mathbb{C}$. This will be verified in Steps 4 and 5.

Since g is continuous on $\mathbb{C} \setminus [\Gamma]$, it suffices to prove that g_0 can be extended to a continuous function on D. Referring to the integral formula (4.29) for the winding number of Γ, we can rewrite (4.48) in the form

$$g_0(z) = \frac{1}{2\pi \mathrm{i}} \int_\Gamma \frac{f(w) - f(z)}{w - z}\, dw, \qquad z \in D_0. \tag{4.51}$$

The function h defined on $D \times D$ by

$$h(w, z) := \begin{cases} \dfrac{f(w) - f(z)}{w - z} & \text{if} \quad w \neq z \\ f'(z) & \text{if} \quad w = z, \end{cases} \qquad w, z \in D,$$

is continuous on $D \times D$. This is obvious at all points (w, z) with $w \neq z$, so it only remains to consider continuity of h at points $(z_1, z_1) \in D \times D$. Fixing $z_1 \in D$ and denoting by D_1 a disk centered at z_1 which is contained in D, we see that the line segment $[z, w]$ lies in D_1 (and hence in D) for all $z, w \in D_1$. Since f is analytic, it follows from the fundamental theorem of calculus (Theorem 4.2.11) that h can be represented as an integral along this segment

$$h(w, z) = \frac{f(w) - f(z)}{w - z} = \frac{1}{w - z} \int_{[z,w]} f'(\zeta)\, d\zeta = \int_0^1 f'\big(z + t(w - z)\big)\, dt,$$

and the standard integral estimate yields

$$|h(w, z) - h(z_1, z_1)| \leq \max_{t \in [0,1]} \big| f'\big(z + t(w - z)\big) - f'\big(z_1\big) \big|, \qquad z, w \in D_1.$$

Since f' is continuous, the right-hand side tends to zero as $(w, z) \to (z_1, z_1)$, which proves the continuity of h on $D \times D$.

5. In the next step we extend the function g_0 analytically to all of D by setting

$$g_0(z) := \frac{1}{2\pi \mathrm{i}} \int_\Gamma h(w,z)\, dw, \qquad z \in D. \tag{4.52}$$

In order to prove that g_0 is continuous on D, we pick a point z_1 in D and denote by K a closed disk centered at z_1 which is contained in D. Then, for all $z \in K$,

$$|g_0(z) - g_0(z_1)| = \frac{1}{2\pi} \left| \int_\Gamma h(w,z) - h(w,z_1)\, dw \right| \le \frac{L(\Gamma)}{2\pi} \max_{w \in [\Gamma]} |h(w,z) - h(w,z_1)|.$$

Since h is uniformly continuous on $[\Gamma] \times K$, the right-hand side tends to zero as $z \to z_1$, which proves the continuity of g_0 (and hence of g) on D. So, according to the remark at the beginning of Step 4, the function g has an analytic extension from $\mathbb{C} \setminus \Gamma$ to the entire complex plane $\mathbb{C}$.

6. Finally, the standard estimate for the Cauchy integral defining the function g_1, and the fact that $g_2 = 0$ on $\mathbb{C} \setminus D \subset E$, yield that g tends to zero at infinity. Consequently, by Liouville's theorem, $g = 0$ on $\mathbb{C}$. □

Remark 4.3.5. The approximation of the original cycle Γ by a polygonal cycle was needed in Step 4 in order to meet the assumptions of Corollary 4.2.23. Another reason is that the standard integral estimates in Step 4 and Step 5 cannot be applied to arbitrary cycles which are not piecewise smooth (or rectifiable, at least).

Cauchy Integral Theorem Revisited. As one expects, the Cauchy integral formula (4.44) for the derivatives of f can also be extended to the more general situation described in Theorem 4.3.4: if Γ is a null-homologous cycle in the open set $D \subset \mathbb{C}$ and f is analytic in D, then

$$\frac{n!}{2\pi \mathrm{i}} \int_\Gamma \frac{f(z)}{(z - z_0)^{(n+1)}}\, dz = \mathrm{wind}(\Gamma, z_0) \cdot f^{(n)}(z_0), \qquad z_0 \in D \setminus [\Gamma]. \tag{4.53}$$

Furthermore, applying Theorem 4.47 to the analytic function $g(z) := (z - z_0)\, f(z)$ we get the following global version of the Cauchy integral theorem.

Theorem 4.3.6 (Cauchy Integral Theorem III). *Let Γ be a null-homologous cycle in the open set $D \subset \mathbb{C}$ and assume that f is analytic in D. Then the integral of f along Γ vanishes,*

$$\int_\Gamma f(z)\, dz = 0.$$

Cauchy Integrals and Phase Portraits. Figure 4.16 shows an analytic function f in a square D and its Cauchy integral along the figure-of-eight loop γ depicted on the left. The set $D \setminus [\gamma]$ consists of three connected components. Corresponding to the winding numbers of γ, which are 0 for the exterior domain, 1 inside the left loop, and -1 inside the right loop, the value of the Cauchy integral is equal to 0, f, and $-f$ in the respective domains.

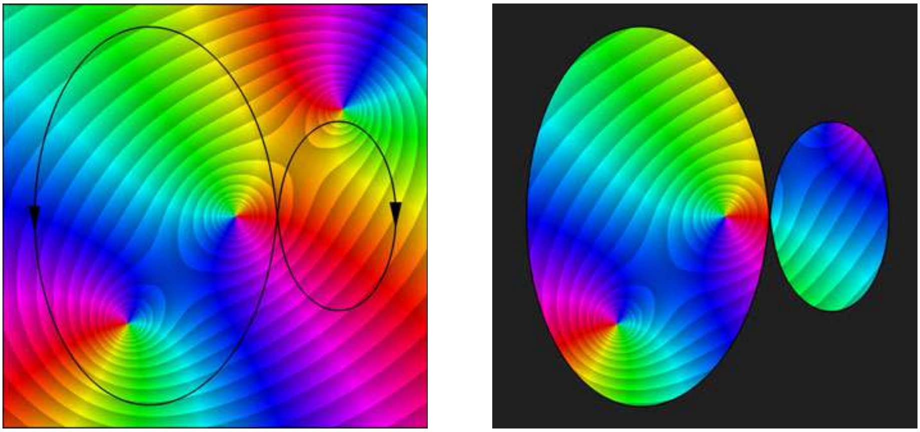


Figure 4.16: An analytic function f and its Cauchy integral along a path

The image in Figure 4.17 (left) shows the phase portrait of the Cauchy integral of the same function for another path γ. Here the winding numbers of γ are 0, 1, and 2, so that the values of the integral in the corresponding domains are 0, f and $2\,f$.

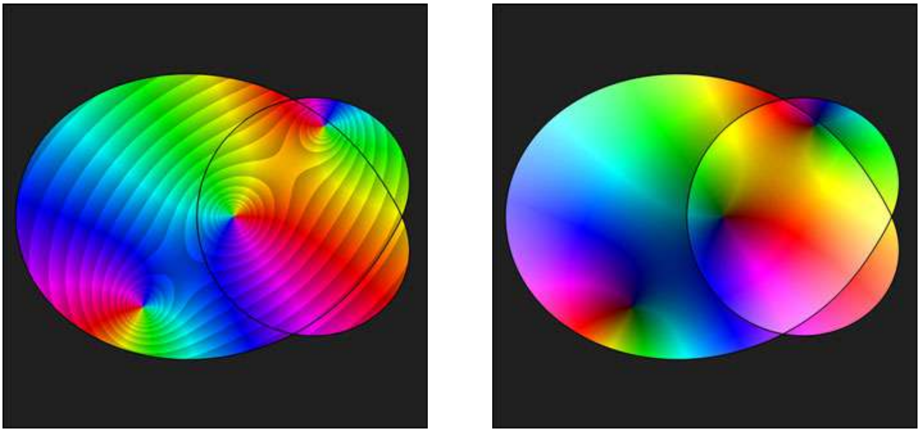


Figure 4.17: Phase portrait and domain coloring of a Cauchy integral

Since the pure phase portraits of f and $2f$ are identical, we have chosen the enhanced version with contour lines. But even then one must inspect the image carefully to detect that something happens across the trace of γ which separates the domain of f from the domain of $2f$ (the contour lines do not fit). The difference between f and $2f$ can be seen a little better in the window on the right, where we visualize the function using domain coloring. But even then one must look attentively to see the slight color nuances which are caused by the factor of 2.

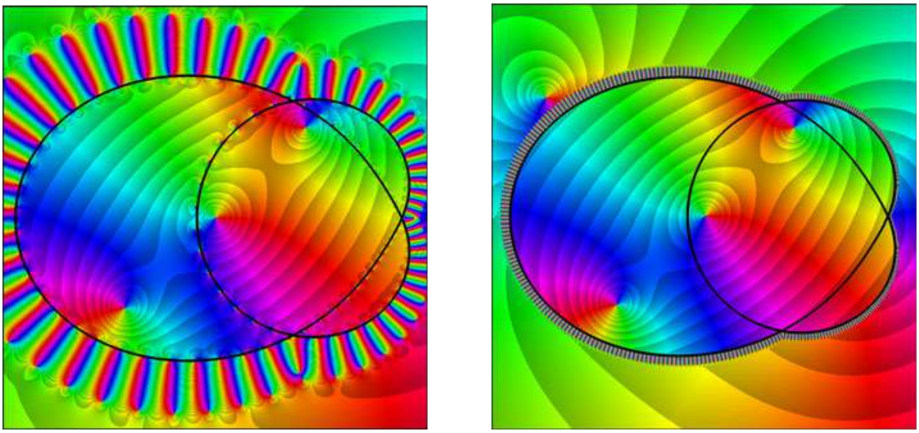


Figure 4.18: Phase portraits of numerically evaluated Cauchy integrals

Numerical Experiments. We remark that the phase portraits in Figures 4.16 and 4.17 have been generated without evaluating the Cauchy integral formula. If the Cauchy integral is computed numerically, we get a completely different picture. Figure 4.18 shows the results of some simple numerical experiments, where the integral is approximated using the trapezoidal rule with 200 (left) respectively 1200 (right) nodes t_k which are uniformly distributed in the parameter interval. The approximating functions depicted in the phase portraits are rational, with poles located at the points $\gamma(t_k)$. It can clearly be seen that the approximation is rather poor near $[\gamma]$. The high density of contour lines in the exterior of the domain encircled by γ reflects the fast decay of the functions in these regions.

4.4 Laurent Series and Singularities

Generalizing the concept of power series by admitting positive and negative powers of $z - z_0$ leads to doubly infinite function series

$$\sum_{k=-\infty}^{\infty} c_k \, (z - z_0)^k := \sum_{k=0}^{\infty} c_k \, (z - z_0)^k + \sum_{k=1}^{\infty} c_{-k} \, (z - z_0)^{-k}. \tag{4.54}$$

Series of this form are called *Laurent series* with center z_0. The first sum on the right-hand side is said to be the *regular part*, while the second sum is called the *main part* (or *principal part* or *singular part*) of the Laurent series (4.54).

Convergence of Laurent Series. Since the regular part of a Laurent series is a power series it converges (absolutely) in a disk. Similarly, the main part converges (absolutely) in the exterior of a closed disk. Note that both sets may be empty or consist of the entire plane. If the regular part converges for $|z - z_0| < R_1$, and the main part for $|z - z_0| > R_0$, with $R_0 < R_1$, then the Laurent series converges in the *ring domain*

$$D := \{z \in \mathbb{C} : 0 \le R_0 < |z - z_0| < R_1 \le \infty\} \tag{4.55}$$

and its sum is an analytic function in D. Conversely, any such function can be expanded into a Laurent series.

Theorem 4.4.1. *Any function $f : D \to \mathbb{C}$ which is analytic in a ring domain (4.55) can be uniquely represented as (the sum of) a Laurent series (4.54) which converges (absolutely) in D. The* Laurent coefficients c_k *are given by the formula*

$$c_k = \frac{1}{2\pi \mathrm{i}} \int_\gamma \frac{f(z)}{(z - z_0)^{k+1}}\, dz, \qquad k \in \mathbb{Z}, \tag{4.56}$$

where integration is along any path γ in D which has winding number one about z_0. Moreover, the Laurent series converges uniformly on any compact subset of D.

Proof. 1. We fix r_0 and r_1 with $R_0 < r_0 < r_1 < R_1$ and denote by γ_0 and γ_1 the positively oriented standard parameterizations of the circles with center z_0 and radii r_0 and r_1, respectively. The cycle $\Gamma := \gamma_1 + \gamma_0^-$ is null homologous in D. Since $\operatorname{wind}(\Gamma, z) = 1$, for all z in the ring domain $D_0 := \{z \in \mathbb{C} : r_0 < |z - z_0| < r_1\}$, the homology version of Cauchy's integral formula Theorem 4.3.4 tells us that

$$f(z) = \frac{1}{2\pi \mathrm{i}} \int_\Gamma \frac{f(w)}{w - z}\, dw = \frac{1}{2\pi \mathrm{i}} \int_{\gamma_1} \frac{f(w)}{w - z}\, dw - \frac{1}{2\pi \mathrm{i}} \int_{\gamma_0} \frac{f(w)}{w - z}\, dw. \tag{4.57}$$

2. Now we expand the Cauchy kernel for $w \in \gamma_1$ and $|z - z_0| < r_1$ into the convergent geometric series

$$\frac{1}{w - z} = \frac{1}{w - z_0} + \frac{z - z_0}{(w - z_0)^2} + \frac{(z - z_0)^2}{(w - z_0)^3} + \dots,$$

and for $w \in \gamma_0$ and $|z - z_0| > r_0$ into the convergent geometric series

$$-\frac{1}{w - z} = \frac{1}{z - z_0} + \frac{w - z_0}{(z - z_0)^2} + \frac{(w - z_0)^2}{(z - z_0)^3} + \dots.$$

Since both series converge uniformly with respect to w (for fixed z), we can insert these representations into (4.57) and interchange summation and integration (for details see the proof of Theorem 4.2.20), which yields the representation

$$f(z) = \sum_{k=-\infty}^{\infty} c_k\, (z - z_0)^k \tag{4.58}$$

of $f(z)$ for $r_0 < |z - z_0| < r_1$. Here the coefficients c_k are given by (4.56) with $\gamma = \gamma_0$ for $k < 0$ and $\gamma = \gamma_1$ for $k \ge 0$. Because f is analytic in D, and since γ_0 and

γ_1 are homotopic to the path γ in Theorem 4.4.1, both paths can be replaced by γ in the integral representation of c_k. Finally, r_0 and r_1 can be chosen arbitrarily close to R_0 and R_1, respectively, so that the Laurent expansion holds for all $z \in D$.

3. In order to prove uniform convergence of the Laurent series on compact subsets of D, we fix a radius r with $R_0 < r < R_1$, and let γ be the (positively oriented) standard parametrization of the circle with radius r and center z_0. If M denotes the maximum of $|f|$ on γ, then the standard integral estimate yields that the Laurent coefficients (4.56) satisfy

$$|c_k| \leq \frac{M}{r^k}, \qquad k \in \mathbb{Z}. \tag{4.59}$$

By the Cauchy–Hadamard criterion (Theorem 3.2.1), the regular part of the Laurent series (4.54) converges absolutely for all $z \in \mathbb{C}$ with $|z - z_0| < r$, while the main part converges absolutely for all $z \in \mathbb{C}$ with $|z - z_0| > r$. Since r can be chosen arbitrarily in (R_0, R_1), *both* series converge (absolutely) in D. Moreover, the inequalities (4.59) guarantee that convergence is uniform in any ring

$$\{z \in \mathbb{C} : R_0 < r_0 \leq |z - z_0| \leq r_1 < R_1\},$$

and hence on any compact subset of D. □

Laurent Decomposition. The main part and the regular part of the Laurent series define two functions f_- and f_+, analytic in $|z - z_0| > R_0$ and $|z - z_0| < R_1$, respectively. Moreover, f_- is analytic at infinity and $f_-(\infty) = 0$. The (unique) representation $f = f_- + f_+$ is said to be the *Laurent decomposition* of f.

Example 4.4.1. Figure 4.19 shows the phase portrait of the function f given by the Laurent series

$$f(z) = \sum_{k=1}^{\infty} \frac{1}{2^{k^2}} \left(z^{k^2} + z^{-k^2} \right), \tag{4.60}$$

which converges in the ring domain $\{z \in \mathbb{C} : 1/2 < |z| < 2\}$. Its Laurent decomposition has the form

$$f(z) = f_+(z) + f_-(z),$$

where f_+ is defined by

$$f_+(z) = \sum_{k=1}^{\infty} \left(\frac{z}{2}\right)^{k^2}, \quad |z| < 2, \tag{4.61}$$

and, in the special case at hand, $f_-(z) = 1/f_+(z)$. The phase portraits of the functions f_- and f_+ are depicted in Figure 4.20.

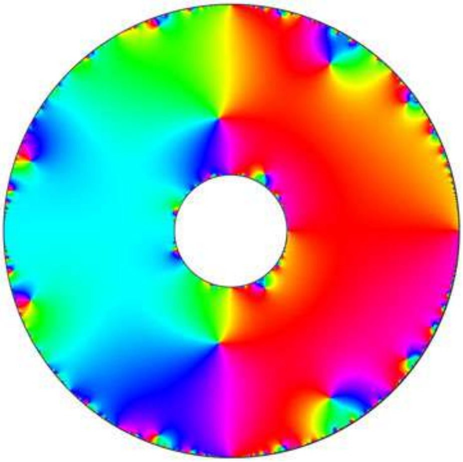


Figure 4.19: Phase portrait of f in (4.60)

In the phase portraits of Figures 4.19 and 4.20 we observe rather irregular behavior near the boundaries of the domains. This happens because the series (4.61) is a so-called *Hadamard gap series.* These power series are characterized by the property that there are sufficiently large gaps between the non-zero terms $c_k\,(z-z_0)^k$. More precisely, the orders k with $c_k \neq 0$ form a sequence $k_1, k_2, \dots$ such that $k_{j+1} \geq q\,k_j$ for all j and some $q > 1$. The *Ostrowski-Hadamard gap theorem* says that the disk of convergence K of a Hadamard gap series is the natural domain of analyticity for its sum f, i.e., f cannot be extended to an analytic function in a domain larger than K (see [32], pp. 119-120).

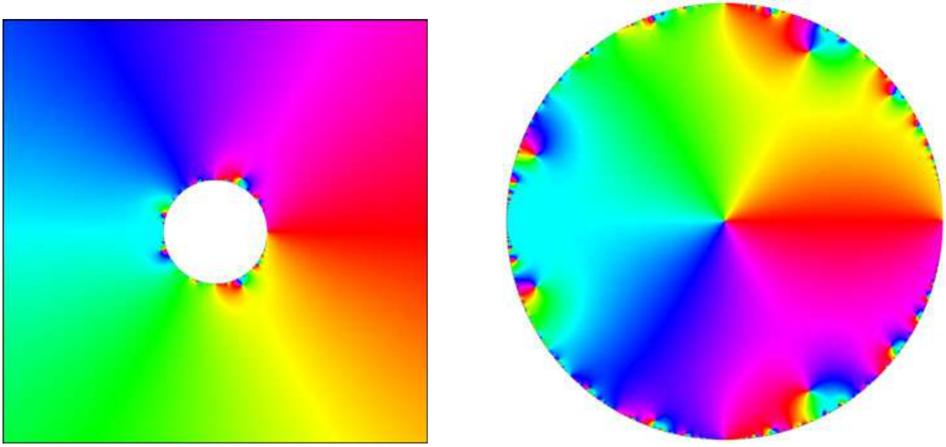


Figure 4.20: The functions f_+ and f_- in the Laurent decomposition of f

Isolated Singularities. The case where f is analytic in a punctured disk $\dot{D}$ is of special interest when we study *isolated singularities* of analytic functions f.

Definition 4.4.2. A point z_0 is said to be an *isolated singularity* of a function f, if there exists a disk D centered at z_0 such that f is analytic in the punctured disk $\dot{D} = D \setminus \{z_0\}$, but not in the full disk D.

In Definition 4.4.2 we also allow z_0 to be the point at infinity, in which case D is understood as a 'disk' on the Riemann sphere $\widehat{\mathbb{C}}$. For the time being we assume that $z_0 \in \mathbb{C}$ and $f : \dot{D} \to \mathbb{C}$, more general situations will be considered later.

If f is analytic in $\dot{D}$, there are two cases when z_0 can be an isolated singularity of f – either f is not defined at z_0, or it is discontinuous at z_0. In the third case, where f is analytic in $\dot{D}$ and continuous at z_0, it is automatically analytic in $\mathbb{D}$ (see Corollary 4.2.23).

Classification of Isolated Singularities. The behavior of a function near an isolated singularity $z_0 \in \mathbb{C}$ depends on the structure of its Laurent series

$$f(z) = \sum_{k=-\infty}^{\infty} c_k\,(z - z_0)^k. \tag{4.62}$$

Definition 4.4.3. An isolated singularity z_0 of an analytic function f with Laurent series (4.62) is called

(i) a *removable singularity*, if $c_k = 0$ for all $k < 0$,

(ii) a *pole* of order m, if $c_{-m} \neq 0$ and $c_k = 0$ for all $k < -m < 0$,

(iii) an *essential singularity*, if $c_k \neq 0$ for infinitely many negative k.

In other words, a point z_0 is a removable singularity, a pole, or an essential singularity, if and only if the main part of its Laurent series

$$\sum_{k=-\infty}^{-1} c_k \, (z - z_0)^k$$

vanishes, is finite, or is infinite, respectively.

In case (i) of the definition, the limit $\lim_{z\to z_0} f(z)$ exists and is equal to the coefficient c_0 of the Laurent series. Thus, (re-)defining $f(z_0) := c_0$ we get an analytic function in D, which explains the notion 'removable singularity'. In fact there is a much weaker condition which guarantees that an isolated singularity is removable. We state it as the first assertion of the following theorem.

Theorem 4.4.4. *Let* $f : \dot{D} \to \mathbb{C}$ *be analytic in the punctured disk* $\dot{D} := D \setminus \{z_0\}$*. Then* z_0 *is*

(i) *a removable singularity if and only if* f *is bounded in a neighborhood of* z_0*,*

(ii) *a pole if and only if* $\lim_{z\to z_0} f(z) = \infty$*,*

(iii) *an essential singularity if and only if* $\lim_{z\to z_0} f(z)$ *does not exist, either as a finite or as an infinite limit.*

Proof. 1. The 'only-if'-direction of the first statement is obvious. To prove the 'if'-direction we may assume that $|f|$ is bounded by M in all of $\dot{D}$. Choosing a circle of radius r in $\dot{D}$ with center z_0 and referring to the estimate (4.59) we get that

$$|c_k| \le \frac{M}{r^k}, \qquad k \in \mathbb{Z},$$

and letting $r \to 0$ we get $c_k = 0$ for all negative k.

2. If f has a pole of order m at z_0, its Laurent series tells us that f has the representation $f(z) = (z - z_0)^{-m} \, g(z)$, where g is analytic in D and $g(z_0) \neq 0$. Consequently, $\lim_{z\to z_0} f(z) = \infty$.

Conversely, assuming that the limit of f at z_0 is ∞, we may suppose that $f(z) \neq 0$ in $\dot{D}$ (decreasing the radius of this disk, if necessary). Then $g := 1/f$ is analytic in $\dot{D}$ and $\lim_{z\to z_0} g(z) = 0$. Consequently z_0 is a removable singularity of g, and setting $g(z_0) := 0$ extends g to an analytic function in D. The extended function g (which is not identically zero) has a zero at z_0, say of order m. Then $f = 1/g$ has a pole of order m at z_0.

3. The third assertion is a logical consequence of the first and the second statements. □

Typical functions with removable singularities and poles are quotients f/g of analytic functions. For example, $f(z) = (\sin z)/z$ has a removable singularity at $z_0 = 0$. More generally, if f and g have zeros of order m and n at z_0, respectively, then z_0 is a removable singularity of f/g if $n \leq m$ and a pole of order $n - m$ otherwise. The functions $f(z) = \exp(1/z)$ and $f(z) = \exp(1/z^2)$ have an essential singularity at $z_0 = 0$, their phase portraits are shown in Figure 3.21. A more interesting example is considered next.

Example 4.4.2. Let q be a complex number with $0 < |q| < 1$. The Laurent series

$$f(z) = \sum_{k=-\infty}^{\infty} q^{k^2} z^k,$$

which we already met in Example 3.3.3 in connection with the Jacobi Theta functions, converges for all $z \in \mathbb{C} \setminus \{0\}$. The function f has an essential singularity at $z_0 = 0$.

Figure 4.21 shows phase portraits of f for $q = 0.9$ and $q = 0.97 \exp(\mathrm{i}\pi/6)$. The black line is the unit circle. It is depicted here since f is invariant with respect to the substitution $z \mapsto 1/z$, which causes a kind of symmetry. In fact the function f obeys yet another symmetry property, namely

$$f(z) = q\, z\, f(q^2\, z), \qquad z \in \dot{\mathbb{C}}. \tag{4.63}$$

This functional equation can easily be verified and shows that, up to multiplication by the factor qz, the function f is invariant with respect to the similarity transformation $z \mapsto q^2 z$. In particular this observation explains the spiral-like pattern of zeros in the right-hand picture: if z_0 is a zero of f, then so is $q^2\, z_0$.

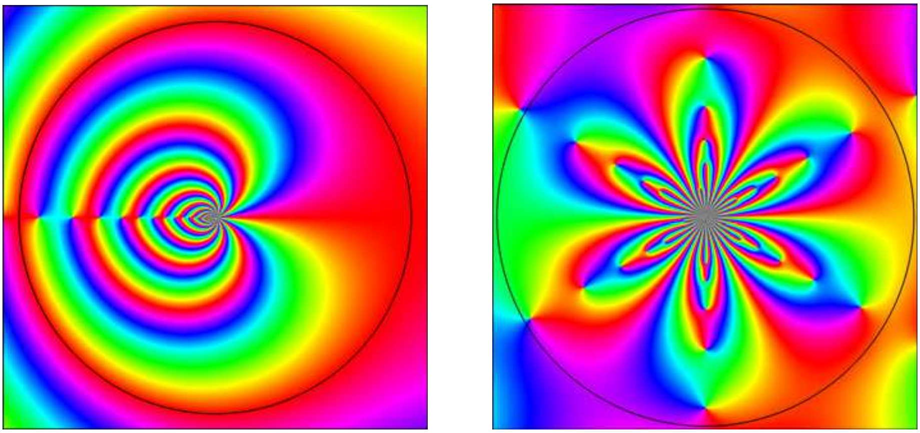


Figure 4.21: Phase portraits of the functions f_q with $q = 0.9$ and $q = 0.97\, \mathrm{e}^{\mathrm{i}\pi/6}$

Behavior Near Isolated Singularities. The phase portraits in Figures 3.21 and 4.21 indicate a rather wild behavior of these functions in a neighborhood of essential singularities. The following theorem shows that this must always be so.

Theorem 4.4.5 (Casorati-Weierstrass). *If z_0 is an essential singularity of the analytic function f in the punctured disk $\dot{D} = D \setminus \{z_0\}$, then $f(\dot{D})$ is dense in the complex plane.*

Since $\dot{D}$ can be replaced by a punctured disk with arbitrarily small radius, the conclusion of the theorem can be rephrased: for any $\varepsilon > 0$ and any $a \in \mathbb{C}$ there exist $z \in \dot{D}$ with $|z - z_0| < \varepsilon$ such that $|f(z) - a| < \varepsilon$.

Proof. Suppose that $f(\dot{D})$ is not dense in $\mathbb{C}$, i.e., there exist $\varepsilon > 0$ and $a \in \mathbb{C}$ such that $|f(z)-a| \geq \varepsilon$ for all $z \in \dot{D}$. Then the function g defined by $g(z) := 1/(f(z)-a)$ is analytic and bounded in $\dot{D}$, so that it has a removable singularity at z_0. We extend g to an analytic function in all of D, represent f as $f(z) = a + 1/g(z)$ for $z \in \dot{D}$, and see that z_0 is either a removable singularity (if $g(z_0) \neq 0$) or a pole (if $g(z_0) = 0$) of f. ☐

The Casorati-Weierstrass theorem is amazing, but in fact there is an even stronger result. The *Great Picard Theorem* tells us that any analytic function with an essential singularity at z_0 takes on all possible complex values – with at most a single exception – infinitely often in any neighborhood of z_0 (see Krantz [31]).

Isolated Singularities in Phase Portraits. Let us now investigate the appearance of isolated singularities in the phase portrait. It is clear that removable singularities cannot be seen at all, while poles manifest themselves in the well-known pattern of isochromatic lines emerging from the singularity. But what about essential singularities?

Theorem 4.4.6. *An isolated singularity z_0 of an analytic function f is an essential singularity if and only if any neighborhood of z_0 intersects infinitely many isochromatic lines of the phase portrait of one and the same color.*

Proof. The 'if'-direction is clear, since a sufficiently small circular neighborhood of a removable singularity or a pole intersects only finitely many isochromatic lines of one color.

In order to prove the converse, we assume that f has an essential singularity at z_0 and show that then any punctured disk $\dot{D}$ centered at z_0 contains an arbitrary number of points at which f attains the same value.

Let $\dot{D}$ be a punctured disk with center z_0. To begin with, we choose a point $z_1 \in \dot{D}$ with $f(z_1) \neq 0$ and an open disk U_1 centered at z_1 such that $\overline{U}_1 \subset \dot{D}$ and $0 \notin V_1 := f(U_1)$. By the open mapping principle the set V_1 is open.

In the next step we apply the Casorati-Weierstrass theorem to find a point $z_2 \in \dot{D} \setminus \overline{U}_1$ with $w_2 := f(z_2) \in V_1$. Further we choose an open disk U_2 centered at z_2 such that $\overline{U}_2 \subset \dot{D} \setminus U_1$ and $V_2 := f(U_2) \subset V_1$. Note that V_2 is an open set.

Continuing in this manner, we get a sequence of pairwise disjoint disks $U_1, U_2, \ldots, U_n \subset \dot{D}$ with images $V_k := f(U_k)$ forming a family of nested open sets $V_1 \supset V_2 \supset \ldots \supset V_n$. Since V_n is open and f' has at most a countable number of zeros, we can choose $w^* \in V_n$ such that for $c := w^*/|w^*|$ the isochromatic set $S := \{z \in \dot{D} : f(z)/|f(z)| = c\}$ contains no zero of f'. Consequently S is the union of smooth isochromatic lines on which $|f|$ is strictly monotone (see page 147). Because w^* belongs to V_n, and hence to all V_k, there exist $z_k^* \in U_k$ with

$$f(z_1^*) = f(z_2^*) = \ldots = f(z_n^*) = w^*.$$

This implies that all points $z_1, \ldots, z_n$ must lie on different isochromatic lines with color c. Since n can be chosen arbitrarily large this proves our claim. $\square$

So far we have investigated isolated singularities at points $z_0 \in \mathbb{C}$. The case $z_0 = \infty$ can be treated analogously by considering the function $g(z) := f(1/z)$ in a neighborhood of $z_0 = 0$. Any function f which is analytic in a ‘punctured disk’ $\dot{D} := \{z \in \mathbb{C} : |z| > r\}$ has a convergent Laurent series with center at infinity,

$$f(z) = \sum_{k=-\infty}^{\infty} c_k\, z^k, \qquad z \in \dot{D}, \tag{4.64}$$

and the behavior of f in a neighborhood of infinity depends in very much the same way as for finite z_0 on the main part of this series, which is now defined by

$$\sum_{k=1}^{\infty} c_k\, z^k. \tag{4.65}$$

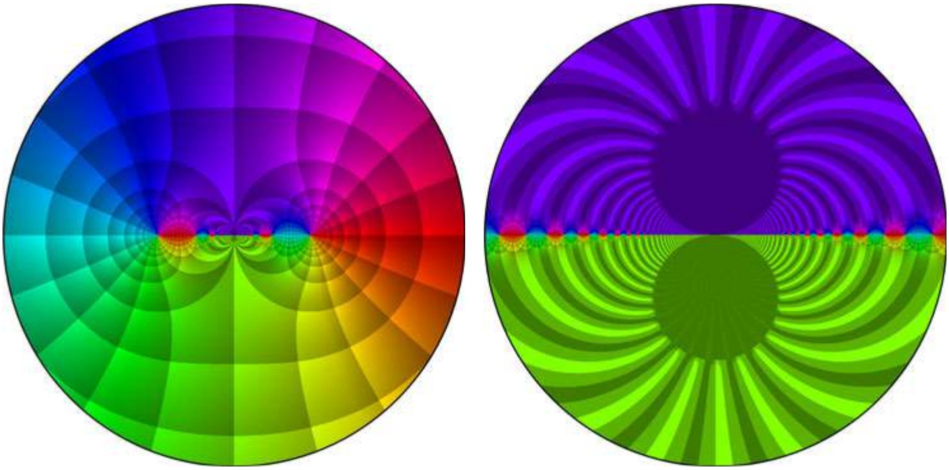


Figure 4.22: The essential singularity of the tangent function at infinity

Simple examples of functions with an essential singularity at infinity are the exponential function and the sine function. More generally, any entire function which is not rational has an essential singularity at infinity. Figure 4.22 shows enhanced phase portraits of the tangent function in a neighborhood of $z_0 = \infty$. The domains depicted are given by $|z| > 1/2$ (left) and $|z| > 10$ (right). The missing structure in the two circular domains in the image on the right is due to the insufficient accuracy of numerical computations. The two pictures demonstrate that essential singularities need not manifest themselves in impressive phase portraits since the wild behavior may be located in tiny regions which are barely visible.

4.5 Residues

It follows from the uniqueness principle that an analytic function in a domain D is completely determined by its Laurent series at an isolated singularity z_0. More surprisingly, much information is encoded just in a *single coefficient* of this series.

Definition 4.5.1. Let $z_0 \in \mathbb{C}$ be an isolated singularity of f. The coefficient c_{-1} of the Laurent series (4.54) is said to be the *residue* of f at z_0,

$$\operatorname{Res}(f, z_0) := c_{-1}.$$

The reason for the somewhat strange name becomes clear when we consider the integral of f along a closed path γ which winds itself once around z_0 in the positive direction. If f is analytic in a punctured disk $\dot{D} := D \backslash \{z_0\}$ and $\gamma \subset \dot{D}$, the formula (4.56) for the Laurent coefficients tells us that

$$\operatorname{Res}(f, z_0) = \frac{1}{2\pi \mathrm{i}} \int_\gamma f(z)\, dz. \tag{4.66}$$

So the residue is essentially 'what remains' of the integral when we integrate around an isolated singularity.

Computation of Residues. It is clear that the residue $\operatorname{Res}(f, z_0)$ at a *removable singularity* of f vanishes. If z_0 is a *simple pole* of f, the residue can easily be computed using the formula

$$\operatorname{Res}(f, z_0) = \lim_{z \to z_0} (z - z_0)\, f(z). \tag{4.67}$$

If f has a *pole of order* n at z_0, the function $g(z) := (z - z_0)^n f(z)$ has a removable singularity at z_0, and its derivative of order $n-1$ satisfies $g^{(n-1)}(z_0) = (n-1)!\, c_{-1}$, which yields

$$\operatorname{Res}(f, z_0) = \frac{1}{(n-1)!} \lim_{z \to z_0} \frac{d^{n-1}}{dz^{n-1}} (z - z_0)^n\, f(z). \tag{4.68}$$

Example 4.5.1. The function $f(z) = z/(\mathrm{e}^z - 1)$ has isolated singularities at the points $z_k = 2k\pi\mathrm{i}$, $k \in \mathbb{Z}$. The singularity at the origin is removable, the other singularities are simple poles. Using (4.67) and de L'Hospital's rule we obtain for $k = \pm 1, \pm 2, \ldots$,

$$\begin{aligned} \operatorname{Res}(f, z_k) &= \lim_{z\to z_k} (z - z_k)\, f(z) \\ &= \lim_{z\to z_k} \frac{(z - z_k)z}{\mathrm{e}^z - 1} \\ &= \lim_{z\to z_k} \frac{2z - z_k}{\mathrm{e}^z} \\ &= 2k\pi\mathrm{i}. \end{aligned}$$

Figure 4.23 shows a phase portrait of f in the square

$$\{z \in \mathbb{C} : |\operatorname{Re} z| < 10,\ |\operatorname{Im} z| < 10\}.$$

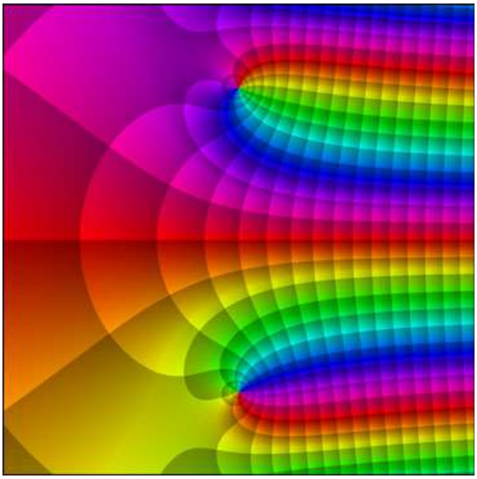


Figure 4.23: $f(z) = z/(\mathrm{e}^z - 1)$

Example 4.5.2. The function $f(z) = \sin z/(1 - \cos z^3)$ has an infinite number of isolated singularities, among them a pole of order 5 at $z_0 = 0$. Figure 4.24 shows a phase portrait of f in the square $|\operatorname{Re} z| < 2$, $|\operatorname{Im} z| < 2$. An attempt to compute the residue $\operatorname{Res}(f, 0)$ using formula (4.68) with $n = 5$ is rather discouraging. In such cases it is often more convenient to work directly with the Laurent series of the functions involved. Using the rules for manipulating power series (see Section 3.2), we get successively

$$\begin{aligned} \sin z &= z - \frac{z^3}{3!} + \frac{z^5}{5!} + O(z^7) \\ \cos z^3 &= 1 - \frac{z^6}{2!} + O(z^{12}) \\ \frac{1}{1 - \cos z^3} &= \frac{2}{z^6} + O(1) \\ \frac{\sin z}{1 - \cos z^3} &= \frac{2}{z^5} - \frac{1}{3\,z^3} + \frac{1}{60\,z} + O(z). \end{aligned}$$

Consequently $\operatorname{Res}(f, 0) = 1/60$.

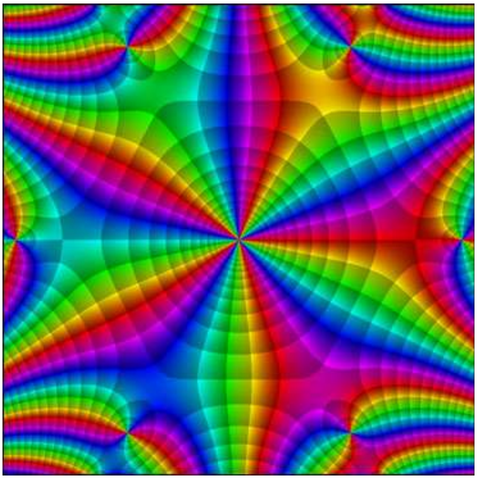


Figure 4.24: $f(z) = \sin z/(1 - \cos z^3)$

If z_0 is an *essential singularity*, computing the Laurent series is usually the easiest way of finding the residue.

Example 4.5.3. The function $f(z) = (z-1)\cos(1/z)$ has an essential singularity at the origin. Figure 4.25 shows a phase portrait of f in the square $|\mathrm{Re}\, z| < 1/2$, $|\mathrm{Im}\, z| < 1/2$.

In order to evaluate the residue $\mathrm{Res}\,(f,0)$ we start with the Taylor series of the cosine function,

$$\cos z = 1 - \frac{z^2}{2!} + \frac{z^4}{4!} \mp \dots .$$

Substituting z by $1/z$ we get the Laurent series

$$\cos\frac{1}{z} = 1 - \frac{1}{2!\, z^2} + \frac{1}{4!\, z^4} \mp \dots$$

of $\cos 1/z$ at the origin, and finally

$$f(z) = z - 1 - \frac{1}{2\, z} + \frac{1}{2\, z^2} \mp \dots$$

is the Laurent series of f at $z_0 = 0$. Consequently $\mathrm{Res}\,(f,0) = -1/2$.

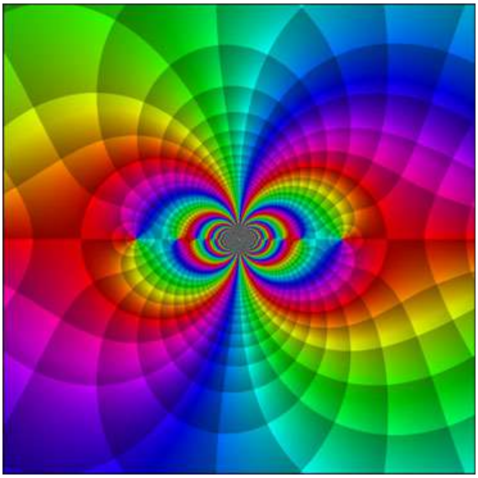


Figure 4.25: $f(z) = (z-1)\cos(1/z)$

Residues at Infinity. The residue $\mathrm{Res}\,(f,\infty)$ of a function f with an isolated singularity *at infinity* is defined by

$$\mathrm{Res}\,(f,\infty) := -c_{-1}, \tag{4.69}$$

where c_{-1} denotes the coefficient of $1/z$ in the Laurent series (4.64) of f at infinity. Instead of calculating residues at ∞ directly from the Laurent series, one can make use of the formula

$$\mathrm{Res}\,(f,\infty) = \mathrm{Res}\,(g,0), \quad \text{where} \quad g(z) := -f(1/z)/z^2.$$

If f is analytic in $D := \{z \in \mathbb{C} : |z| > r\}$, $r > 0$, and γ is a path in D with winding number *minus one* about zero (which is another way of saying that γ winds once around the point at infinity in the positive direction), then

$$\mathrm{Res}\,(f,\infty) = -c_{-1} = \frac{1}{2\pi \mathrm{i}} \int_\gamma f(z)\, dz. \tag{4.70}$$

Note that the residue $\mathrm{Res}\,(f,z_0)$ for finite z_0 is involved in the *principal part* of the Laurent series, while $\mathrm{Res}\,(f,\infty)$ appears in the *regular part*. So the integral in (4.70) may be different from zero even if infinity is a removable singularity of f.

The Residue Theorem. The next result is an extended version of Cauchy's integral theorem in the presence of isolated singularities.

Theorem 4.5.2 (Residue Theorem)**.** *Let D be an open set in the plane, assume that $S := \{z_1, \ldots, z_n\}$ is a finite subset of D, and let $f : D \setminus S \to \mathbb{C}$ be analytic. Then, for any cycle Γ in $D \setminus S$ which is null-homologous in D,*

$$\int_\Gamma f(z)\, dz = 2\pi \mathrm{i} \sum_{k=1}^{n} \operatorname{wind}(\Gamma, z_k) \cdot \operatorname{Res}(f, z_k).$$

Proof. In the first step we complement Γ by adding paths which encircle the singular points of f to a cycle $\widetilde{\Gamma}$ which is null-homologous in $D \setminus S$.

Since all points z_k are interior points of D, we find pairwise disjoint closed disks with centers z_k and sufficiently small radii r_k which are contained in the set $(D \setminus S) \cup \{z_k\}$. For $k = 1, \ldots, n$ we define the paths $\gamma_k : [0, 1] \to D$ by

$$\gamma_k(t) := r_k \exp(-2n_k\pi\, \mathrm{i}\, t), \quad \text{where} \quad n_k := \operatorname{wind}(\Gamma, z_k).$$

Then $\widetilde{\Gamma} := \Gamma + \gamma_1 + \ldots + \gamma_n$ is a null-homologous cycle in $D \setminus S$. Figure 4.26 illustrates this situation. The dark-blue circle in the right picture is run through twice.

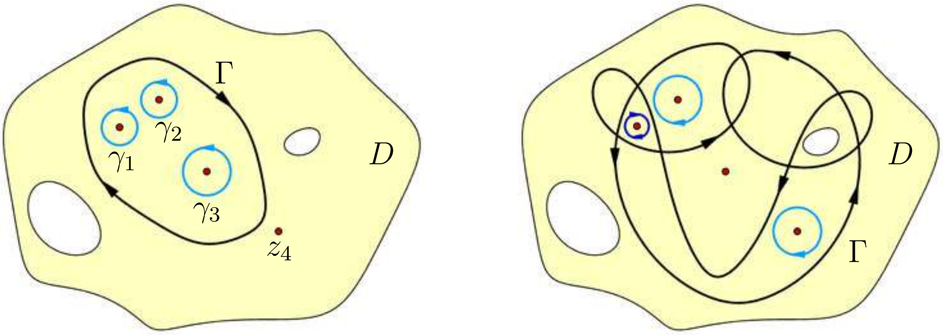


Figure 4.26: Complementing a cycle Γ to obtain a null-homologous cycle

Indeed, for all points $z \in \mathbb{C} \setminus (D \cup S)$ we have $\operatorname{wind}(\gamma_k, z) = 0$ (since γ_k is null-homotopic in $\mathbb{C} \setminus \{z\}$ to the point z_k), so that $\operatorname{wind}\big(\widetilde{\Gamma}, z\big) = \operatorname{wind}\big(\Gamma, z\big) = 0$. If $z = z_j \in S$, then

$$\operatorname{wind}(\gamma_k, z_j) = \begin{cases} 0 & \text{if} \quad k \neq j \\ -n_j & \text{if} \quad k = j, \end{cases}$$

and we have again

$$\operatorname{wind}\big(\widetilde{\Gamma}, z_j\big) = \operatorname{wind}\big(\Gamma, z_j\big) + \operatorname{wind}(\gamma_j, z_j) = n_j - n_j = 0.$$

Since the integral of f along γ_j is just $n_j \operatorname{Res}(f, z_j)$, the assertion now follows from the general form of Cauchy's integral theorem (Theorem 4.3.6. □

On applying Theorem 4.5.2 to a positively oriented circular path which contains all singularities, and invoking formula (4.70) for the residue $\operatorname{Res}(f,\infty)$, we get the following corollary.

Corollary 4.5.3. *If f is analytic in the entire plane with the exception of finitely many isolated singularities $z_1,\dots,z_n$, then*

$$\operatorname{Res}(f,z_1)+\ldots+\operatorname{Res}(f,z_n)+\operatorname{Res}(f,\infty)=0.$$

Applications. The residue theorem is not only useful in complex analysis, it has vast applications in real analysis as well. In particular it is a powerful tool for evaluating real integrals. We demonstrate this for a specific class of integrals described in the following theorem.

Theorem 4.5.4. *Let P and Q be polynomials in two variables x and y, assume that $Q(x,y)\neq 0$ whenever $x,y\in\mathbb{R}$ and $x^2+y^2=1$, and let $R:=P/Q$. Then*

$$\int_0^{2\pi} R(\cos t,\sin t)\,dt = 2\pi\sum_{k=1}^{n}\operatorname{Res}(f,z_k), \tag{4.71}$$

where

$$f(z):=\frac{1}{z}\,R\left(\frac{1}{2}\left(z+\frac{1}{z}\right),\frac{1}{2\mathrm{i}}\left(z-\frac{1}{z}\right)\right),\qquad z\in\mathbb{C}, \tag{4.72}$$

and $z_1,\dots,z_n$ are the poles of f in the unit disk $\mathbb{D}$.

Proof. First of all we remark that f is a rational function of the complex variable z. If $z:=x+\mathrm{i}y\in\mathbb{T}$, we have $x=(1/2)(z+1/z)$ and $y=(1/2\mathrm{i})(z-1/z)$, so that $f(z)=R(x,y)/(x+\mathrm{i}y)$ with $x^2+y^2=1$. Consequently f has no poles on the unit circle.

By definition of the integral along the positively oriented unit circle $\mathbb{T}$ with the parametrization $\gamma:[0,2\pi]\to\mathbb{T}, t\mapsto \mathrm{e}^{\mathrm{i}t}$ we get

$$\int_\gamma f(z)\,dz=\int_0^{2\pi}\mathrm{e}^{-\mathrm{i}t}R\big(\tfrac{1}{2}(\mathrm{e}^{\mathrm{i}t}+\mathrm{e}^{-\mathrm{i}t}),\tfrac{1}{2\mathrm{i}}\,(\mathrm{e}^{\mathrm{i}t}-\mathrm{e}^{-\mathrm{i}t})\big)\,\mathrm{i}\mathrm{e}^{\mathrm{i}t}\,dt=\mathrm{i}\textstyle\int_0^{2\pi}R(\cos t,\sin t)\,dt.$$

Evaluating the integral on the left-hand side by the residue theorem yields the desired result. □

The following example demonstrates that formula (4.71) may still hold even when Q does not satisfy all assumptions of Theorem 4.5.4. The reason is a partial cancellation of zeros in the numerator and the denominator of R after the substitution (4.72).

Example 4.5.4. We would like to evaluate the integral

$$\int_0^{2\pi}\frac{\sin^4 t}{1+\cos t}\,dt,$$

which has the form (4.71) with $R(x,y) := y^4/(1+x)$. Here one assumption of Theorem 4.5.4 is violated since the denominator $Q(x,y) := 1+x$ has a zero satisfying $x^2+y^2=1$, namely $x=-1$, $y=0$. This zero of Q corresponds to a (double) zero of the denominator $1+\cos t$ of the integrand $g(t) := \sin^4 t/(1+\cos t)$, but in fact it cancels with the fourth order zero of the numerator of g at the same point $t=\pi$. Setting $g(\pi) := 0$ makes g a continuous function on $[0,2\pi]$.

Though this cancellation effect cannot be observed in the function $R(x,y)$, it becomes obvious after substituting x and y according to (4.72), which yields the function

$$f(z) := \frac{1}{z}\,\frac{\left(\frac{1}{2\mathrm{i}}\left(z-\frac{1}{z}\right)\right)^4}{1+\frac{1}{2}\left(z+\frac{1}{z}\right)} = \frac{1}{2^3 z^4}\,\frac{(z^2-1)^4}{(z^2+2z+1)} = \frac{1}{8}\,\frac{(z-1)^4(z+1)^2}{z^4}. \qquad (4.73)$$

The only pole of this rational function at $z_1=0$ has order four. In order to determine the residue of f at z_1 we thus need the coefficient of z^3 of the polynomial $(z-1)^4(z+1)^2$. This coefficient equals 4, so that $\operatorname{Res}(f,0) = 4/8 = 1/2$, and finally

$$\int_0^{2\pi} \frac{\sin^4 t}{1+\cos t}\,dt = 2\pi \operatorname{Res}(f,0) = \pi.$$

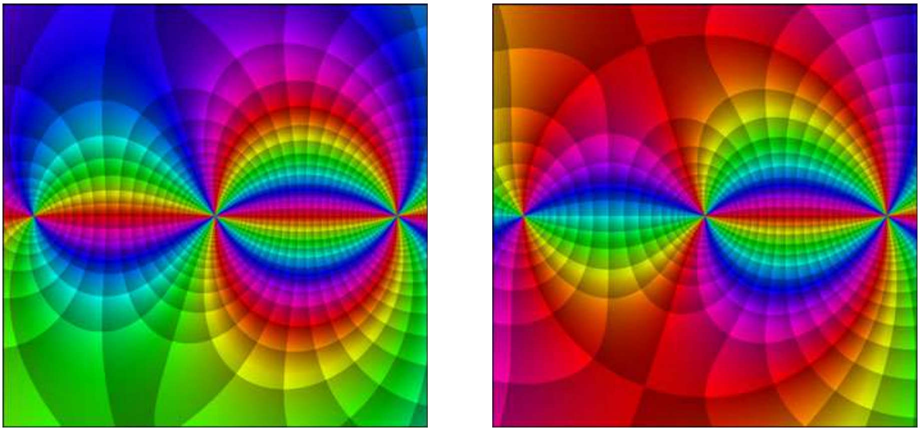


Figure 4.27: Phase portraits of the functions f and $z \mapsto z\,f(z)$

Figure 4.27 shows phase portraits of the functions f and $h : z \mapsto z\,f(z)$ in the square defined by $|\operatorname{Re} z| < 1.2$, $|\operatorname{Im} z| < 1.2$. The zeros of f and h at ± 1 and the poles at the origin can clearly be identified in these pictures. The (real) values of the integrand g are associated to the values of h along the unit circle by $g(t) = h(\cos t + \mathrm{i}\sin t) = R(\cos t, \sin t)$ with $t \in [0,2\pi]$.

The next example illustrates an application of the residue theorem to the evaluation of an *improper integral.*

Example 4.5.5. The *Fourier transform* F of the function $f(t) = 1/(1+t^2)$ is given by the (absolutely convergent) improper integral

$$F(\omega) := \int_{-\infty}^{+\infty} \frac{\mathrm{e}^{\mathrm{i}\omega t}}{1+t^2}\, dt, \qquad \omega \in \mathbb{R}. \tag{4.74}$$

In order to evaluate the integral by the residue theorem, we 'complexify' the integrand and replace the integral (4.74) by an integral along an appropriately chosen closed contour.

The asymptotic behavior of the function $g(z) := \mathrm{e}^{\mathrm{i}\omega z}/(1+z^2)$ in the complex plane depends on the parameter ω. If $\omega > 0$ it decays (exponentially) in the upper half-plane as $\operatorname{Im} z \to +\infty$; if $\omega < 0$ we have exponential decay in the lower half-plane. This can be seen in the phase portraits of g for $\omega = 5$ (left) and $\omega = -5$ (right) of Figure 4.28. We remark that one can also read off the monotone decay of f along the real axis from the inclination of the isochromatic lines against the vertical direction.

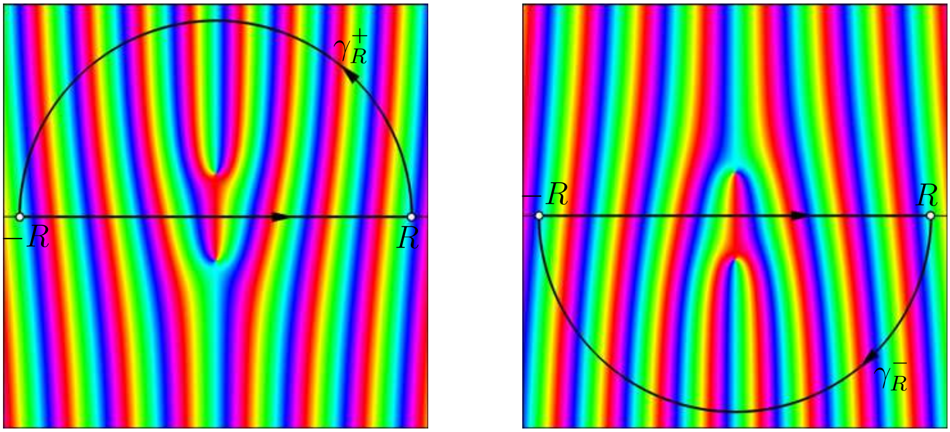


Figure 4.28: The function $g(z) := \mathrm{e}^{\mathrm{i}\omega z}/(1+z^2)$ for $\omega = 5$ and $\omega = -5$

Assuming that ω is non-negative, we consider a closed path γ_R which consists of the segment $[-R, R]$ on the real axis and a (positively oriented) semi-circle γ_R^+ in the upper half plane with radius $R > 1$ and center at the origin. We evaluate the integral of g along γ_R by the residue theorem, and prove that the contribution of γ_R^+ tends to zero as R goes to infinity, so that in the limit only the integral along the real axis remains.

The function g is analytic in the plane with the exception of two simple poles at $z = \pm\mathrm{i}$. Since the pole at $-\mathrm{i}$ lies in the exterior of $[\gamma_R]$ we only need the residue of g at i,

$$\operatorname{Res}(g, \mathrm{i}) = \lim_{z\to\mathrm{i}} (z - \mathrm{i})\, \frac{\mathrm{e}^{\mathrm{i}\omega z}}{1+z^2} = \frac{\mathrm{e}^{-\omega}}{2\mathrm{i}}.$$

The residue theorem tells us that for all $R > 1$

$$\int_{-R}^{R} g(z)\, dz + \int_{\gamma_R^+} g(z)\, dz = \int_{\gamma_R} g(z)\, dz = 2\pi\mathrm{i}\,\mathrm{Res}\,(g, \mathrm{i}) = \pi\,\mathrm{e}^{-\omega}. \tag{4.75}$$

In order to estimate the integral of g along γ_R^+ we observe that for all $\omega \geq 0$ and $z \in \gamma_R^+$

$$\left|\mathrm{e}^{\mathrm{i}\omega z}\right| \leq 1, \qquad \left|z^2 + 1\right| \geq R^2 - 1,$$

so that the standard integral estimate yields

$$\left|\int_{\gamma_R^+} g(z)\, dz\right| \leq \frac{\pi\, R}{R^2 - 1}.$$

Since the right-hand side tends to zero as $R \to \infty$, we infer from (4.75) that

$$\int_{-\infty}^{+\infty} \frac{\mathrm{e}^{\mathrm{i}\omega z}}{1 + z^2}\, dz = \pi\,\mathrm{e}^{-\omega}, \qquad \omega \geq 0. \tag{4.76}$$

If $\omega < 0$ we replace the semi-circle γ_R^+ in the upper half-plane by the corresponding semi-circle γ_R^- in the lower half-plane, which is now run through in the clockwise direction. Using $\mathrm{Res}\,(g, -\mathrm{i}) = \mathrm{i}\,\mathrm{e}^{\omega}/2$, we obtain similarly after some calculation

$$\int_{-\infty}^{+\infty} \frac{\mathrm{e}^{\mathrm{i}\omega z}}{1 + z^2}\, dz = \pi\,\mathrm{e}^{\omega}, \qquad \omega < 0. \tag{4.77}$$

Combining formulas (4.76) and (4.77) we finally arrive at

$$F(\omega) := \int_{-\infty}^{+\infty} \frac{\mathrm{e}^{\mathrm{i}\omega z}}{1 + z^2}\, dz = \pi\,\mathrm{e}^{-|\omega|}, \qquad \omega \in \mathbb{R}.$$

Further applications of the residue theorem to integral transformations will be given in Volume 2.

The Logarithmic Residue. We close this section with an alternative approach to the *argument principle* (Theorem 3.5.9) which is provided by the residue theorem. Our starting point is the *logarithmic derivative* f'/f of an analytic function f. If z_0 is a zero of order $n > 0$ or a pole of order $-n > 0$, then f can be represented as $f(z) = (z - z_0)^n\, g(z)$, where g is analytic in a neighborhood of z_0 and $g(z_0) \neq 0$. A short calculation yields that for all z in a punctured neighborhood of z_0

$$\frac{f'(z)}{f(z)} = \frac{n}{z - z_0} + \frac{g'(z)}{g(z)}, \tag{4.78}$$

so that z_0 is a simple pole of f'/f with residue n. The number n is therefore called the *logarithmic residue* of f at z_0.

The two windows on the left-hand side of Figure 4.29 show phase portraits in a neighborhood of a zero of order three and a pole of order five. The two windows to the right depict the corresponding logarithmic derivatives.

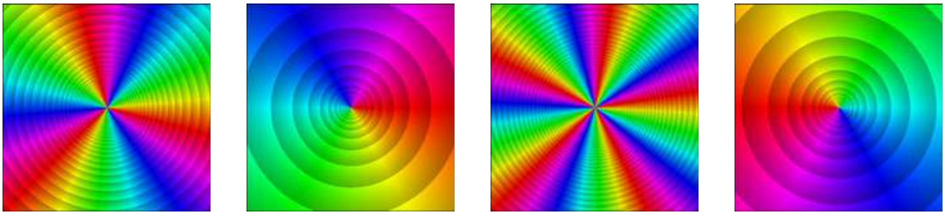


Figure 4.29: A triple zero and a pole of order five with the logarithmic derivatives

Let f be *meromorphic* in the simply connected domain D, and assume that J is a positively oriented Jordan curve in D which meets neither zeros nor poles of f. Then the residue theorem tells us that

$$\int_J \frac{f'(z)}{f(z)}\, dz = 2\pi \mathrm{i} \sum_j \operatorname{Res}\,(f'/f, z_j) = 2\pi \mathrm{i}\,(N-P), \tag{4.79}$$

where N and P denote the numbers (counted with multiplicity) of zeros and of poles of f which are located in the interior of J. The integral on the left-hand side can easily be evaluated: if $\gamma : [0,1] \to J$ is a parametrization of J, then any continuous branch $\log(f \circ \gamma)$ of the logarithm is a primitive of f'/f along γ, so that

$$\int_J \frac{f'(z)}{f(z)}\, dz = \log(f \circ \gamma)(1) - \log(f \circ \gamma)(0) = 2\pi \mathrm{i}\, \mathrm{wind}_J f. \tag{4.80}$$

Combining (4.79) and (4.80) we again obtain the argument principle (Theorem 3.5.9).

4.6 Conjugate Harmonic Functions

In this section we investigate another remarkable consequence of the Cauchy–Riemann equations: the real part u and the imaginary part v of an analytic function satisfy a special partial differential equation of second order.

Definition 4.6.1. A function $u : D \to \mathbb{R}$ is said to be *harmonic* in the open set $D \subset \mathbb{R}^2$ if it is twice continuously differentiable and satisfies the *Laplace equation*

$$\Delta u := \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0 \tag{4.81}$$

in D. The differential operator on the left-hand side is called the *Laplace operator*.

The Laplace equation is certainly one of the most important partial differential equations, and harmonic functions are ubiquitous in almost all fields of mathematical physics and many other applications. The Laplace operator and the concept of harmonic functions have far reaching extensions to higher dimensions and to functions on manifolds.

Analytic and Harmonic Functions. In the following we identify subsets of $\mathbb{C}$ with the corresponding subsets of the Euclidean plane $\mathbb{R}^2$.

Theorem 4.6.2. *If $f : D \to \mathbb{C}$ is analytic in an open set $D \subset \mathbb{C}$, then $u := \operatorname{Re} f$ and $v := \operatorname{Im} f$ are harmonic in D.*

Proof. Since f is infinitely differentiable, we can differentiate the first Cauchy-Riemann equation with respect to x and the second equation with respect to y,

$$\frac{\partial^2 u}{\partial x^2} = \frac{\partial^2 v}{\partial x \partial y}, \qquad \frac{\partial^2 u}{\partial y^2} = -\frac{\partial^2 v}{\partial y \partial x}.$$

Interchanging the order of differentiation and adding these equations yields the desired result. ☐

Conjugate Harmonic Functions. By Theorem 4.6.2, any analytic function generates *two* real harmonic functions u and v which are *coupled by the Cauchy-Riemann equations*. Such functions are said to be *conjugate harmonic*. More precisely, we shall say that v is conjugate harmonic to u, but because v is conjugate harmonic to u if and only if u is conjugate harmonic to $-v$, the relation is often symmetrized by neglecting the sign.

Contour Lines. The contour lines

$$\{(x, y) \in \mathbb{R}^2 : u(x, y) = \text{const}\}$$
$$\{(x, y) \in \mathbb{R}^2 : v(x, y) = \text{const}\}$$

of conjugate harmonic functions u and v have an interesting property. Figure 4.30 shows a number of such lines, where u and v are the real part (red lines) and the imaginary part (blue lines) of the polynomial

$$f(z) = z^3 + z - 1.$$

The domain depicted is the square $|x| < 1$, $|y| < 1$. The next theorem confirms the expectation that the mutual orthogonality of the families of red and blue lines is not an accident.

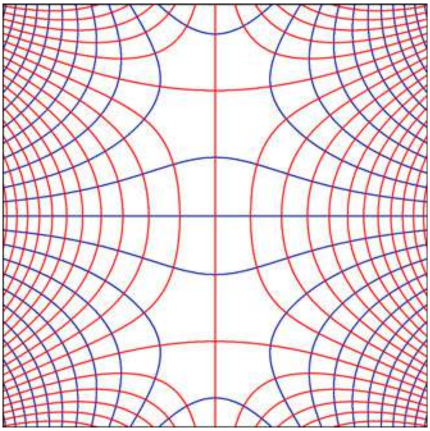


Figure 4.30: Contour lines of conjugate harmonic functions

Theorem 4.6.3. *If u and v are conjugate harmonic and have non-vanishing gradients, then their contour lines are mutually orthogonal.*

Proof. The Cauchy–Riemann equations tell us that the gradient of u and the gradient of v are orthogonal,

$$\left(\frac{\partial u}{\partial x}, \frac{\partial u}{\partial y}\right) \cdot \left(\frac{\partial v}{\partial x}, \frac{\partial v}{\partial y}\right) = \left(\frac{\partial u}{\partial x}, \frac{\partial u}{\partial y}\right) \cdot \left(-\frac{\partial u}{\partial y}, \frac{\partial u}{\partial x}\right) = 0.$$

Since both gradients do not vanish and are orthogonal to the corresponding contour lines, the result follows. □

The functions $u(x, y) = x$ and $v(x, y) = 2y$ illustrate that the converse does not hold, that is, orthogonality of the contour lines of harmonic functions u and v does not necessarily imply that u and v are *conjugate* harmonic. Nevertheless, there is an ‘almost converse’.

Theorem 4.6.4. *Let u and v be real-valued harmonic functions with non-vanishing gradient in a domain D. If the contour lines of u and v are mutually orthogonal, then there exists a real constant c such that cv is conjugate harmonic to u, i.e., $u + \mathrm{i}\, cv$ is analytic in D.*

Proof. By assumption, u and v are twice continuously differentiable, the gradients of u and v are orthogonal and non-zero. Consequently, there exists a continuously differentiable non-vanishing function $c : D \to \mathbb{R}$ such that

$$\frac{\partial u}{\partial x} + \mathrm{i}\, \frac{\partial u}{\partial y} = -c\, \mathrm{i} \left(\frac{\partial v}{\partial x} + \mathrm{i}\, \frac{\partial v}{\partial y}\right).$$

Separating real and imaginary parts we get

$$\frac{\partial u}{\partial x} = c\, \frac{\partial v}{\partial y}, \qquad \frac{\partial u}{\partial y} = -c\, \frac{\partial v}{\partial x}. \tag{4.82}$$

Differentiating the first equation with respect to x, the second with respect to y, and adding the results we get

$$0 = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial^2 y} = \frac{\partial c}{\partial x}\, \frac{\partial v}{\partial y} - \frac{\partial c}{\partial y}\, \frac{\partial v}{\partial x}, \tag{4.83}$$

which tells us that the gradients of c and v are orthogonal in D. Multiplying the equations in (4.82) by $1/c$, differentiating the first equation with respect to y, the second with respect to x, and subtracting the results we obtain similarly that the gradient of $1/c$ is orthogonal to the gradient of u. Since the gradients of u and v span $\mathbb{R}^2$, and the gradients of c and $1/c$ are parallel, the gradient of c vanishes in D. Because D is connected, c is a constant, and (4.82) are the Cauchy-Riemann equations for $u + \mathrm{i}\, cv$. □

The next result gives a first answer to the question: which harmonic functions have conjugates?

Theorem 4.6.5. *Any harmonic function $u : D \to \mathbb{R}$ in a* simply connected *domain D has a harmonic conjugate v in D. The function v is uniquely determined up to an additive constant.*

Proof. 1. If $D_0 \subset D$ is a disk centered at z_0, we define v in D_0 by the integral

$$v(z) := -\int_\gamma \frac{\partial u}{\partial y}(z)\,dx + \int_\gamma \frac{\partial u}{\partial x}(z)\,dy, \quad z \in D_0, \tag{4.84}$$

where γ is a path from z_0 to z in D_0. Since u is harmonic, the integrability condition

$$-\frac{\partial}{\partial y}\frac{\partial u}{\partial y} = \frac{\partial}{\partial x}\frac{\partial u}{\partial x}$$

is fulfilled, so that the integral does not depend on the choice of the path from z_0 to z. Moreover, v is continuously differentiable and

$$\frac{\partial v}{\partial x} = -\frac{\partial u}{\partial y}, \qquad \frac{\partial v}{\partial y} = \frac{\partial u}{\partial x}.$$

So the Cauchy-Riemann equations are satisfied, and because $f_0 := u + \mathrm{i}\,v$ is continuously differentiable, it is analytic in D_0. Since the Cauchy–Riemann equations determine the gradient of the conjugate function v, any two such functions can differ at most by a constant.

2. If D_0 and D_1 are two overlapping disks in D, the corresponding analytic functions f_0 and f_1 constructed in the first step can be adjusted so that $f_0 = f_1$ in $D_0 \cap D_1$. It follows that f_0 has an unrestricted analytic continuation in D, and by the monodromy principle (Theorem 3.6.10) it generates an analytic function f in D. Clearly, the imaginary part of f is a harmonic conjugate of u. ☐

Corollary 4.6.6. *Harmonic functions in an open set $D \subset \mathbb{R}^2$ are infinitely differentiable in D.*

Proof. If u is harmonic in D and $D_0 \subset D$ is a disk, then u has a harmonic conjugate v in D_0. Since $f = u + \mathrm{i}v$ is analytic, u and v are infinitely differentiable in D_0. ☐

Another consequence of Theorem 4.6.5 is the *mean value property* of harmonic functions.

Theorem 4.6.7. *If u is harmonic in a disk D with center z_0 and radius R, then for any $r < R$*

$$u(z_0) = \frac{1}{2\pi}\int_0^{2\pi} u\big(z_0 + r\mathrm{e}^{\mathrm{i}t}\big)\,dt.$$

For a proof we complement u by a conjugate harmonic function v to an analytic function $u + \mathrm{i}v$ in D and apply the mean value theorem for analytic functions (Corollary 4.2.3).

The function $u(x, y) = \log|x + \mathrm{i}\, y|$ shows that a global harmonic conjugate need not always exist. This function is harmonic in $\mathbb{C} \setminus \{0\}$, because it is (locally) the real part of (any branch of) $\log z$. Assuming that there exists a harmonic conjugate v of u in $\mathbb{C} \setminus \{0\}$, we can normalize it so that $v(1) = 0$. Then $u + \mathrm{i}\, v$ is analytic in $\mathbb{C}\setminus\{0\}$ and coincides with the principal branch $\operatorname{Log} z$ in a neighborhood of $z_0 = 1$. Consequently $u+\mathrm{i}\, v$ is an analytic continuation of $\operatorname{Log} z$ to all of $\mathbb{C}\setminus\{0\}$, but this function is discontinuous at the negative real axis.

Conjugate harmonic functions play an important role in a number of applications, in particular in problems of plane electrostatics and hydrodynamics.

Plane Electrostatics. The electric field $E = (E_x, E_y)$ in a region which contains no charges is *irrotational*, $\operatorname{rot} E = 0$, and has *zero divergence*, $\operatorname{div} E = 0$. In two dimensions these conditions are expressed by

$$\frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} = 0, \qquad \frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} = 0, \tag{4.85}$$

i.e., the electric field is governed by the Cauchy–Riemann equations for the analytic function $f := E_x - \mathrm{i}E_y$. However, this *complex electric field* is seldom used to model problems of (plane) electrostatics directly, it is more usual to work with the *electrostatic potential*, which is a real-valued scalar function Φ whose (negative) gradient is the electric field, $E = -\operatorname{grad}\Phi$. The local existence of an electrostatic potential is guaranteed by $\operatorname{rot} E = 0$, while $\operatorname{div} E = 0$ implies that Φ is harmonic,

$$\Delta\Phi = \operatorname{div}\operatorname{grad}\Phi = -\operatorname{div} E = 0.$$

The contour lines of Φ are called *(equi-)potential lines.* As the negative gradient of Φ, the electric field is everywhere orthogonal to the potential lines. The contour lines of any (local) conjugate harmonic function Ψ of Φ are the *field lines.* The electric field (and hence the force which acts on a test charge) is everywhere tangent to the field lines. The analytic function $F := \Phi + \mathrm{i}\,\Psi$ is said to be a *complex potential* of E. Since

$$E_x = -\frac{\partial\Phi}{\partial x} = -\frac{\partial\Psi}{\partial y}, \qquad E_y = -\frac{\partial\Phi}{\partial y} = \frac{\partial\Psi}{\partial x},$$

the complex electric field $f = E_x - \mathrm{i}E_y$ is the negative derivative of the complex potential, $f = -F'$.

Example 4.6.1. Families of potential lines and field lines give a good idea of the electric field. The electric fields of a single charge (left) and of two positive charges (right), are visualized in the two pictures of Figure 4.31. The field lines are depicted in blue, the potential lines are colored red.

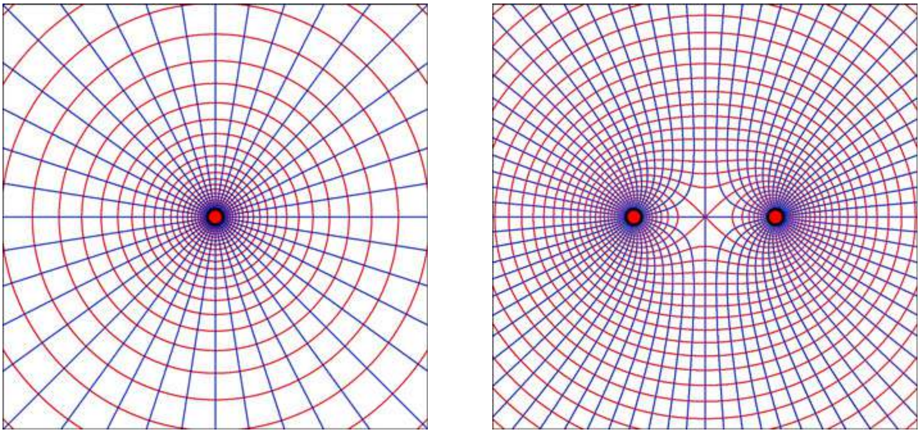


Figure 4.31: The electric fields of two configurations of point charges

Up to physical constants, the potential of a positive unit charge at the origin is $\Phi(z) = -\mathrm{Log}\,|z|$, its potential lines are concentric circles. The potential Φ is the real part of (any branch of the negative of) the complex logarithm

$$F(z) = -\log z = -\mathrm{Log}\,|z| - \mathrm{i}\arg z,$$

and hence $\Psi(z) = -\arg z$ is a (local) harmonic conjugate of Φ. The field lines $\Psi = \text{const}$ are radial rays emerging at the origin.

In contrast to the complex potential $F(z) = -\log z$, which is not globally defined in $\mathbb{C} \setminus \{0\}$, the complex electric field $f(z) = -F'(z) = 1/z$ is a well-defined function on all of $\mathbb{C} \setminus \{0\}$.

We mention that the potential lines $\Phi = \text{const}$ and the field lines $\Psi = \text{const}$ can be read off from the (enhanced) phase portrait of $g := \exp(\Phi + \mathrm{i}\Psi)$: since the equations

$$\Phi = c, \quad \text{and} \quad \mathrm{Log}\,\big|\exp(\Phi + \mathrm{i}\Psi)\big| = c,$$

are equivalent, the modulus contour lines in the enhanced phase portrait of g coincide with the potential lines. Because the isochromatic lines are orthogonal to these lines, they are the field lines.

The potential generated by a configuration of several charges is the sum of their individual potentials. If, for example, negative unit charges are located at the points $z_1, \dots, z_m$ and positive unit charges are sitting at $z_{m+1}, \dots, z_n$, the resulting potential is

$$\Phi(z) = \mathrm{Log}\,|z - z_1| + \ldots + \mathrm{Log}\,|z - z_m| - \mathrm{Log}\,|z - z_{m+1}| - \ldots - \mathrm{Log}\,|z - z_n|.$$

The potential lines are then the modulus contour lines of the rational function

$$g(z) := \exp(\Phi + \mathrm{i}\Psi) = \frac{(z - z_1)\,\cdots\,(z - z_m)}{(z - z_{m+1})\,\cdots\,(z - z_n)},$$

and the field lines are the isochromatic lines in the phase portrait of g. The pictures in Figure 4.32 illustrate this for the case of a negative unit charge at $z = -1$ and a positive unit charge at 1, in which case $g(z) = (z+1)/(z-1)$.

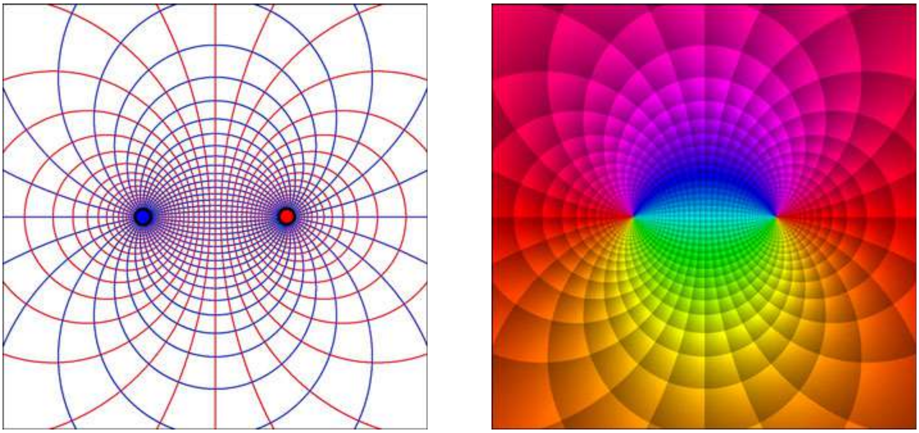


Figure 4.32: Two unit charges with opposite sign sitting at $z = 1$ and $z = -1$

Example 4.6.2 (Electric Dipole). We observe what happens when two opposite charges located at the points $z = -d$ and $z = d$ (with $d \in \mathbb{R}_+$) approach each other at the origin. In order to get a well-defined (and non-trivial) limit, the electric charges will be scaled up by the reciprocal distance $1/(2d)$ of the approaching points. The limit configuration for $d \to 0$ is called a *dipole*.

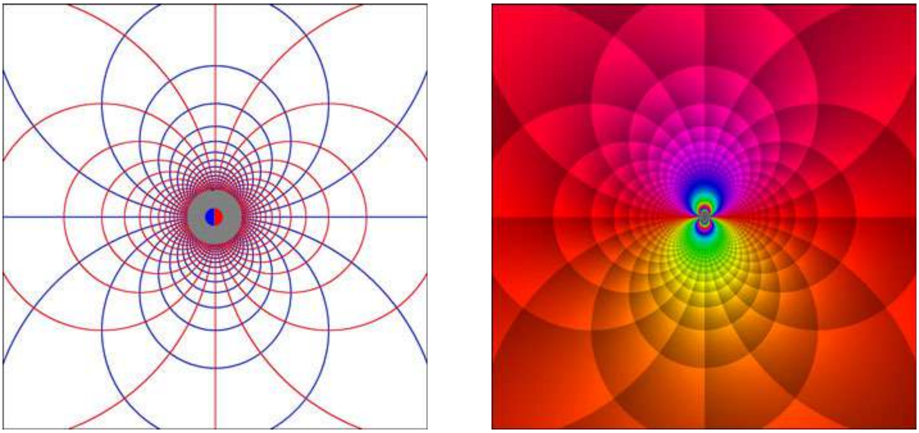


Figure 4.33: The field of a dipole and the phase portrait of $g(z) = \exp(1/z)$

Instead of calculating the dipole potential from its definition,

$$\Phi(z) := \lim_{d \to 0} \frac{1}{2d} \operatorname{Log} \left| \frac{z+d}{z-d} \right|,$$

we determine the limit of the corresponding complex electric field,

$$f(z) = \lim_{d \to 0} \frac{1}{2d} \left(\frac{1}{z-d} - \frac{1}{z+d} \right) = \frac{1}{z^2}.$$

From $F' = -f$ we now get the complex potential $F(z) = \Phi(z) + \mathrm{i}\Psi(z) = 1/z$ of a dipole (the integration constant is of no interest here). Figure 4.33 depicts the electric field of the dipole and the phase portrait of $g(z) = \exp\big(F(z)\big) = \exp(1/z)$. In the gray zone of the left picture the density of lines is too high to depict them individually, the phase portrait on the right has a much better resolution.

Since the electric field E is the negative gradient of the potential Φ, the strength $|E|$ of the field can be read off from the density of the potential lines. It follows from the Cauchy–Riemann equations that

$$|\operatorname{grad} \Phi| = |\operatorname{grad} \Psi|,$$

so that potential lines and field lines have the same density. This relation is reflected in the enhanced phase portrait of the function g by the fact that the cells generated by potential lines and field lines are approximate squares. The edge length of these cells is inversely proportional to the strength of the electric field.

Plane Potential Flow. Another physical problem which is closely related to the Cauchy–Riemann equations and conjugate harmonic functions is two-dimensional *potential flow*. This idealized model of fluid flow is based on the assumptions that the fluid is incompressible and inviscid (without friction), and that its velocity field is time-independent and irrotational.

In order to state the equations of motion, we denote by V_x and V_y the components of the velocity V of the fluid particles in the directions of the (Cartesian) x and y coordinates. Then the velocity field $V = V(x, y)$ in the two-dimensional domain D occupied by the fluid satisfies the partial differential equations

$$\frac{\partial V_x}{\partial x} + \frac{\partial V_y}{\partial y} = 0, \qquad \frac{\partial V_y}{\partial x} - \frac{\partial V_x}{\partial y} = 0. \tag{4.86}$$

The first equation reflects the incompressibility of the fluid (continuity equation, vanishing divergence), while the second one tells us that the flow has vanishing rotation. These equations are just the Cauchy–Riemann equations for the *complex velocity*

$$f(x + \mathrm{i}y) := V_x(x, y) - \mathrm{i}\, V_y(x, y).$$

Consequently, any analytic function in a domain D can be interpreted as the velocity field of a plane potential flow in D, and vice versa.

A (local) primitive F of the complex velocity f is called a *complex velocity potential.* Separating F into its real part Φ and its imaginary part Ψ, and considering Φ and Ψ as functions of x and y, we get

$$V_x = \frac{\partial \Phi}{\partial x} = \frac{\partial \Psi}{\partial y}, \qquad V_y = \frac{\partial \Phi}{\partial y} = -\frac{\partial \Psi}{\partial x}. \tag{4.87}$$

Since the velocity field $V = (V_x, V_y)$ is the gradient of the real function Φ, we refer to Φ as the *real velocity potential.*

As real and imaginary parts of the analytic function F, the functions Φ and Ψ are conjugate harmonic, and hence the contour lines of Ψ are parallel to the gradient of Φ, which is just the velocity V. Hence the equations $\Psi = \text{const}$ characterize the *stream lines*of the flow, and Ψ is designated as *stream function.*

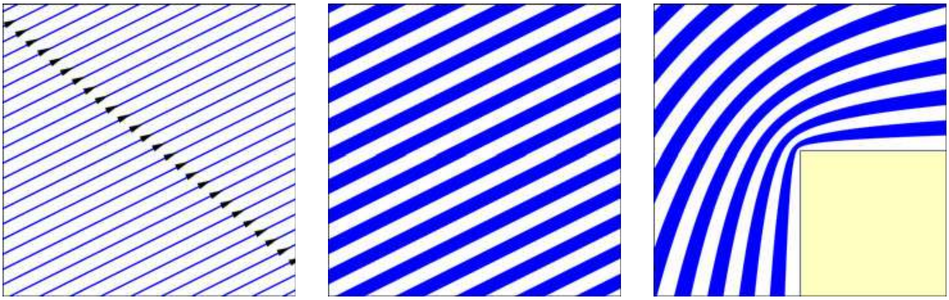


Figure 4.34: Parallel flow and potential flow around a rectangular obstacle

The images on the left and in the middle of Figure 4.34 visualize a *parallel flow* with constant velocity field $V(x, y) = V_0$. The complex velocity and the complex potential of this flow are $f(z) = \overline{V}_0$ and $F(z) = \overline{V}_0\, z$, respectively, where the vector V_0 is interpreted as a complex number. The picture on the left shows the stream lines of the flow. The window in the middle visualizes the same flow, but here we have colored the regions between neighboring stream lines. Since the generation of these stripes is computationally more stable, we shall exclusively use this representation from now on. Interpreting these pictures, one should take into account that the speed of the fluid is inversely proportional to the density of stream lines.

Appropriately chosen branches of the power functions $f(z) = z^a$ describe potential flows around edges. The right window of Figure 4.34 shows the flow past a rectangular corner, which has the complex velocity $f(z) = z^{-1/3}$ and the complex velocity potential $F(z) = (3/2)\, z^{2/3}$. Here the analytic branches of the power functions in the three-quarter plane $\{z \in \mathbb{C} \setminus \{0\} : 0 \leq \arg z \leq 3\pi/2\}$ are chosen such that they are positive on $\mathbb{R}_+$. Since then $\arg f(z) = -\pi/2$ on the negative imaginary axis, the boundary of the obstacle is (the limit of) a stream line of the flow.

We remark that often the complex velocity is easier to handle than the complex velocity potential. For example, the three flows depicted in Figure 4.35 have well-defined complex velocity fields $f(z) = a/z$ with $a \neq 0$ in the punctured plane $\dot{\mathbb{C}} := \mathbb{C} \setminus \{0\}$, while the associated complex potentials $F(z) = a \log z$ are multiple-valued functions.

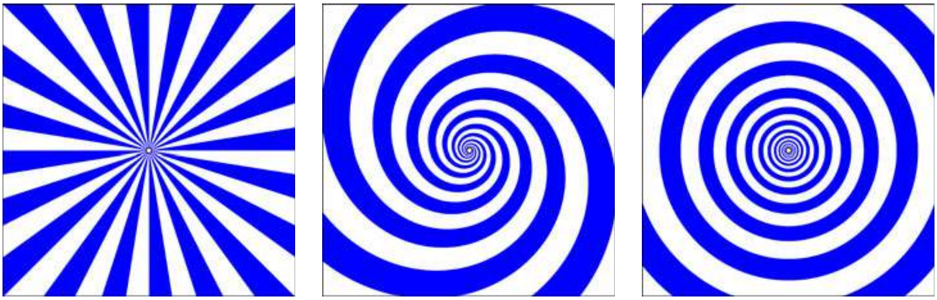


Figure 4.35: Three types of potential flows with a singularity at the origin

All three velocity fields have a singularity at the origin. The window on the left, where $\operatorname{Im} a = 0$, shows a *source* (or a *sink*, depending on the direction of the flow which is determined by the sign of a), the other two depict two types of *vortices*. In the picture in the middle we have $\operatorname{Re} a \neq 0$, $\operatorname{Im} a \neq 0$, while the flow field with closed stream lines on the right side corresponds to a purely imaginary value of a. The latter is a *pure vortex*, also called a *center*.

Another interesting model of an ideal flow is given by the complex *Joukovski potential*,

$$F(z) := v_\infty (z + z^{-1}), \qquad |z| \geq 1, \quad v_\infty \in \mathbb{R}_+. \tag{4.88}$$

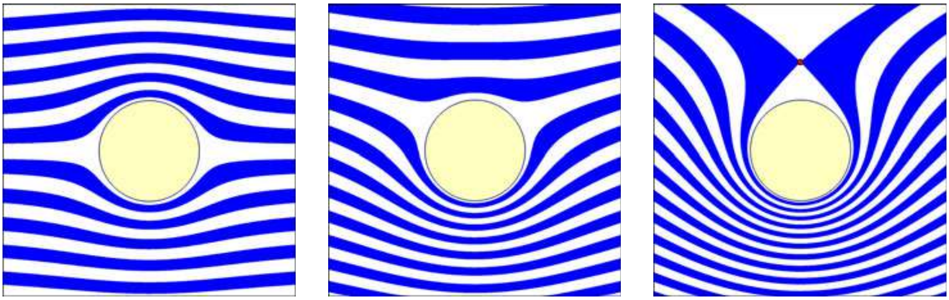


Figure 4.36: Potential flow around a disk (circular cylinder)

It can easily be verified that the unit circle is a contour line of the stream function $\Psi := \operatorname{Im} F$, the potential describes a two-dimensional flow around the unit disk. The complex velocity of the Joukovski flow is $f(z) = v_\infty (1 - z^{-2})$. At large distances from the origin the flow is approximately parallel to the real axis and has speed v_∞. The stream lines are depicted in Figure 4.36 (left).

Since the equations (4.86) of potential flow are linear, velocity fields and potentials can be superimposed. Adding to the Joukovski flow a pure vortex flow centered at the origin, we obtain the complex velocity

$$f(z) = v_\infty \left(1 - \frac{1}{z^2}\right) + \frac{a}{z}, \qquad a \in \mathbb{R}. \tag{4.89}$$

For every real a the velocity field is tangent to the unit circle. The corresponding stream lines for two different non-zero values of a are also shown in Figure 4.36. The picture in the middle corresponds to $0 < a < 2\,v_\infty$, and in the window on the right we have $a > 2\,v_\infty$. In the latter case the flow has a *stagnation point* z_0 (indicated by a red dot), where $f(z_0) = 0$.

Note that the pictures of Figure 4.36 can also be interpreted as cross sections through a three-dimensional potential flow past a circular cylinder, with velocity orthogonal to the cylinder axis.

If $V = (V_x, V_y)$ is the velocity field of a flow and Γ is an oriented curve (path) in the domain D occupied by the fluid, the integrals

$$Q := \int_\Gamma \big(V_x\,dy - V_y\,dx\big) = \int_\Gamma V \cdot N\,ds, \tag{4.90}$$

$$J := \int_\Gamma \big(V_x\,dx + V_y\,dy\big) = \int_\Gamma V \cdot T\,ds, \tag{4.91}$$

define two quantities which are important in applications. In the integrals on the right-hand side, T is a positively oriented tangent vector to the (smooth oriented) curve Γ, the vector $N := -\mathrm{i}\,T$ is perpendicular to Γ, the dot stands for the scalar product, and integration is with respect to arc length. Both quantities admit a physical interpretation: Q is called the *flux* of the vector field V through Γ, and J is referred to as the *circulation* of V along Γ.

In terms of the complex velocity $f = V_x - \mathrm{i}\,V_y$, both equations (4.90) and (4.91) can be summarized in the single relation

$$J + \mathrm{i}\,Q = \int_\Gamma f(z)\,dz. \tag{4.92}$$

By Cauchy's integral theorem, the values of J and Q do not depend on the choice of the curve Γ, provided the curves considered are homotopic in the flow region. In particular, the vortex field

$$f(z) = \frac{Q - \mathrm{i}\,J}{2\pi\,z}$$

has circulation J around and flux Q through any positively oriented Jordan curve which encircles the origin.

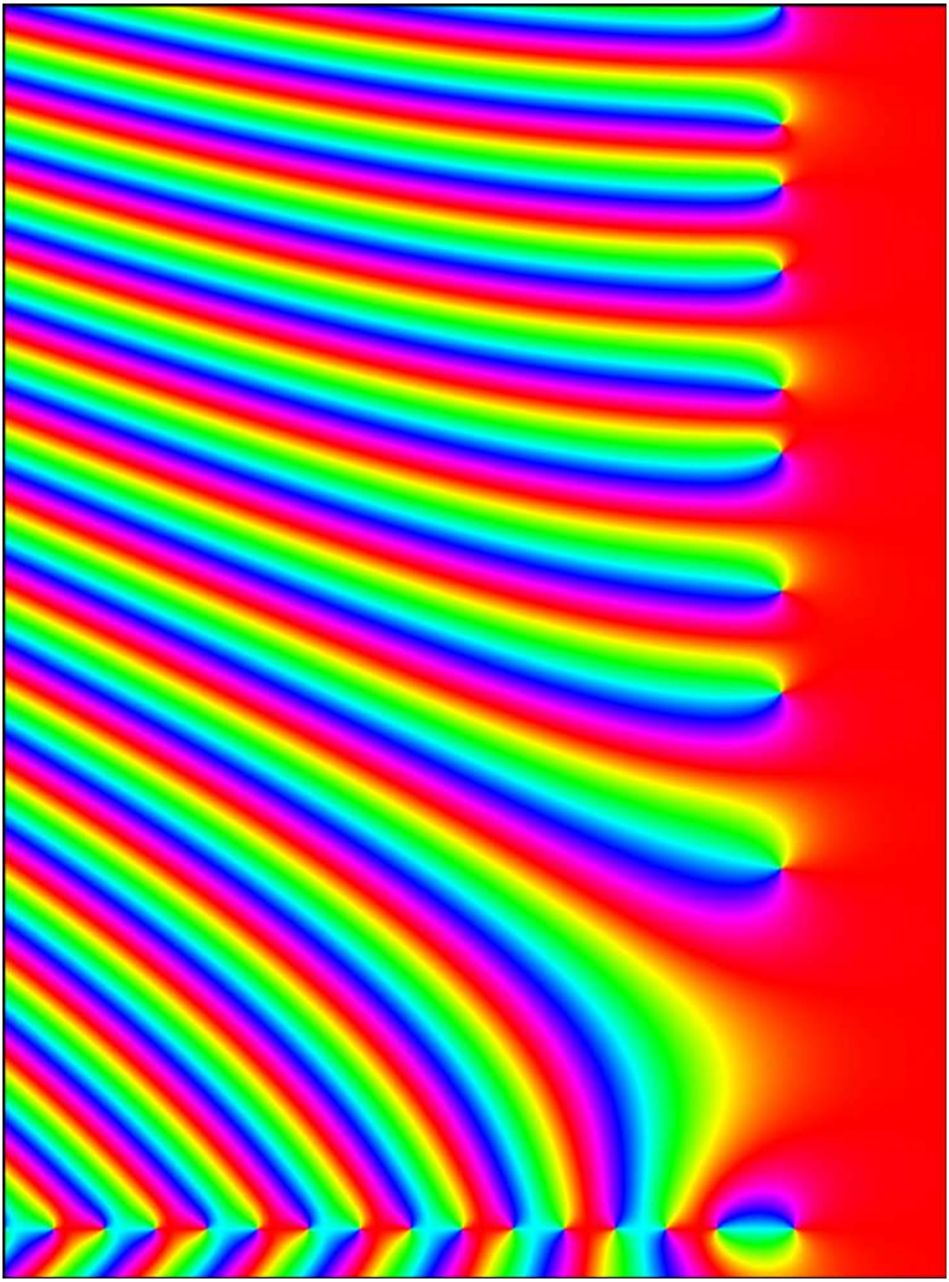


Figure 5.1: The Riemann Zeta function in $-30 < \operatorname{Re} z < 7$, $-2 < \operatorname{Im} z < 48$

Chapter 5

Construction Principles

This chapter is concerned with various methods for constructing analytic functions. We begin by investigating *limit processes* involving analytic functions in Section 5.1. In particular we shall introduce the notion of *normal convergence*, which plays an important role in complex analysis. This concept will be used in Section 5.2, where we prove *Montel's theorem*, a compactness criterion for families of analytic functions. This result is a key ingredient of many existence proofs and will be applied in Section 6.4 to derive the Riemann mapping theorem.

Section 5.3 deals with various types of *function series*. Besides the prominent Taylor, Laurent and (complex) Fourier series, we also study Dirichlet and Lambert series, and we comment briefly on Hadamard gap series.

Infinite products form the subject of Section 5.4. We discuss at length different concepts of convergence and state several convergence criteria. Infinite Blaschke products serve to illustrate the results.

In Section 5.5 we introduce *Cauchy integrals* with arbitrary density and investigate their basic properties. These powerful tools allow us to construct analytic functions with prescribed boundary behavior and will be fundamental to our investigations of boundary value problems in Volume 2.

The last section is devoted to *integrals involving parameters*, another common source of analytic functions. In particular we shall meet such parameter integrals in connection with *integral transforms*, a topic that will be worked out in greater details in Volume 2.

5.1 Function Sequences

In this section we study general *limit procedures* as constructive tools for generating analytic functions. We have already utilized this approach in Section 3.2, where sums of power series were defined as limits of polynomials. In order to gain more flexibility we now extend these procedures to general function sequences.

E. Wegert, *Visual Complex Functions: An Introduction with Phase Portraits*,
DOI 10.1007/978-3-0348-0180-5_5, © Springer Basel 2012

First of all we need a concept of convergence which is appropriate for our purposes. It is clear that pointwise convergence is too weak a property, since it does not even preserve continuity. On the other hand, uniform convergence is too strong. As it happened for power series, we shall often meet sequences which do not converge uniformly on the entire domain set.

Normal Convergence. The most important notion of convergence in the theory of analytic functions lies in between these two extremes. We define it on continuous functions, and do not yet require their analyticity.

Definition 5.1.1. A sequence (f_n) of continuous functions $f_n : X \to Y$ is said to converge *normally* in X to the limit function $f : X \to Y$, if (f_n) converges uniformly to f on each compact subset of X.

In the literature, one finds several alternative names for normal convergence, like "local uniform convergence" or "uniform convergence on compact subsets", or simply "compact convergence". Also since the use of the term "normal convergence" is not absolutely standard in the literature, the reader is advised to be careful when working with it.

Clearly, normal convergence implies pointwise convergence, but the setting of normal convergence has the advantage that it also guarantees continuity of the limit function. And, as we shall see in a minute, normal convergence has much stronger consequences for analytic functions.

A Convergence Criterion. We now consider complex functions defined in open subsets of the complex plane. The following result is often useful in verifying normal convergence.

Lemma 5.1.2. *Let $D \subset \mathbb{C}$ be an open set. Then a sequence of functions f_n defined in D converges normally to a function f in D, if and only if for any $z \in D$ there exists a (closed or open) disk $D_z \subset D$ centered at z such that f_n converges to f uniformly on D_z.*

Proof. For closed disks the necessity of the condition is obvious. On the other hand it is clear that it does not matter whether we consider open or closed disks. Sufficiency then follows, since any compact set $K \subset D$ can be covered by a finite collection of the open disks D_z. $\square$

Convergence of Derivatives. The last lemma will immediately be applied for obtaining the following fundamental result, which has no counterpart in real analysis – normal convergence of analytic functions enforces convergence of their derivatives of any order!

Theorem 5.1.3. *Let (f_n) be a sequence of analytic functions $f_n : D \to \mathbb{C}$ which converges normally on an open set $D \subset \mathbb{C}$ to the limit function f. Then f is analytic in D. Moreover, for every positive integer k, the sequence of the kth derivatives $f_n^{(k)}$ converges normally to $f^{(k)}$.*

Proof. 1. Let K be a closed disk contained in D. Since D is open we can find a larger disk $\widetilde{K}$ that is still contained in D. Denoting by γ the standard parametrization of the boundary of $\widetilde{K}$, and by G the interior of $\widetilde{K}$, we get from Cauchy's integral formula (4.43) that

$$f_n(z_0) = \frac{1}{2\pi \mathrm{i}} \int_\gamma \frac{f_n(z)}{z - z_0}\, dz \tag{5.1}$$

for all z_0 in G. Since (f_n) converges uniformly on the trace of γ we see that f is continuous on $[\gamma]$, and taking the limit in (5.1) using Lemma 4.2.9 we arrive at

$$f(z_0) = \frac{1}{2\pi \mathrm{i}} \int_\gamma \frac{f(z)}{z - z_0}\, dz$$

for all z_0 in G. By Theorem 4.2.20, the limit function f is analytic in G.

2. Now applying Cauchy's integral formula (4.44) for the kth derivatives of f and f_n, we have for all $z_0 \in G$

$$f_n^{(k)}(z_0) - f^{(k)}(z_0) = \frac{k!}{2\pi \mathrm{i}} \int_\gamma \frac{f_n(z) - f(z)}{(z - z_0)^{k+1}}\, dz. \tag{5.2}$$

Because f_n converges uniformly on $[\gamma]$, and since $[\gamma]$ has a positive distance from K, the integrand converges to zero uniformly with respect to $z_0 \in K$, and the standard integral estimate implies uniform convergence of $f_n^{(k)}$ to $f^{(k)}$ on K. ☐

Zeros of the Limit Function. Another advantage of normal convergence is that it preserves locally the number of zeros. A more precise formulation of this vague statement is given in the next theorem.

Theorem 5.1.4 (Hurwitz Theorem)**.** *Let (f_n) be a sequence of analytic functions $f_n : D \to \mathbb{C}$ which converges normally in the open set D to a function f which is not the zero-function. Then, for each point $a \in D$, there exist a disk $D(a) \subset D$ and an integer $N(a)$ such that for all $n \geq N(a)$ the functions f_n and f have the same number of zeros in $D(a)$, counted with multiplicity.*

Proof. By the identity theorem, the zeros of f are isolated. Consequently, for any point $a \in D$, there exists a disk $D(a)$ such that the closure of $D(a)$ is contained in D and contains either no zero of f (if $f(a) \neq 0$) or exactly one zero, namely a. In both cases $|f|$ has a positive minimum $M(a)$ on the (compact) boundary $T(a)$ of $D(a)$.

Since $T(a)$ is compact and f_n converges normally to f, there exists an integer $N(a)$ such that

$$|f_n(z) - f(z)| < M(a) \leq |f(z)|, \qquad z \in T(a),\ n \geq N(a).$$

By Rouché's theorem (Theorem 3.4.7), the functions f_n and f have the same number of zeros in $D(a)$. ☐

Univalent Functions. In the next chapter injective analytic functions will play a special role. In complex analysis it is tradition to endow this class of functions with a special name.

Definition 5.1.5. Analytic functions which are injective are said to be *univalent* (or *schlicht*).

It is of great importance that normal convergence preserves univalence, with one exception.

Corollary 5.1.6. *Let* (f_n) *be a sequence of univalent functions* $f_n : D \to \mathbb{C}$ *which converges normally in the open set* D *to a function* f*. Then the limit function* f *is either constant or univalent in* D*.*

Proof. Assume that f is not univalent. Then there exist two distinct points a and b in D with $f(a) = f(b)$, so that $f - f(a)$ has at least two zeros, namely a and b. The functions $f_n - f(a)$ converge normally to $f - f(a)$. If f is not constant, then $f - f(a)$ is not the zero-function, hence, by Theorem 5.1.4, $f_n - f(a)$ must also have at least two distinct zeros for sufficiently large n. This is impossible because f_n is supposed to be univalent. □

Convergence of the Phase. Using phase portraits for exploring complex functions, it is natural to ask in which way (normal) convergence is reflected in the phase. At the moment we can only give a partial solution to this question, a more satisfactory answer will be obtained in Volume 2 after the necessary tools become available.

Notice that (normal) convergence of a sequence (f_n) does *not* necessarily imply convergence of the phase functions $\psi(f_n)$. For example, the functions $f_n(z) = z^n$ converge normally in the unit disk $\mathbb{D}$ to $f(z) = 0$, but their phase functions do not converge at any point, except at the origin.

Theorem 5.1.7. *Let* $f_n : D \to \mathbb{C}$ *be a sequence of continuous complex functions which converges normally to* f *on* D*, and let* N *denote the zero set of* f*. Then the phase functions* $\psi(f_n)$ *converge normally to* $\psi(f)$ *on* $D \setminus N$*.*

Proof. Let K be a compact subset of $D \setminus N$. Since f is continuous, there exists a positive number C such that $0 < 1/C \le |f| \le C$ on K. Uniform convergence of f_n to f on K implies that $0 < 1/(2C) \le |f_n| \le 2C$ on K for all sufficiently large n. Since the phase ψ is a uniformly continuous function in the ring domain $1/(2C) \le |w| \le 2C$, the assertion follows. □

5.2 Normal Families

In this section we shall prove a simple version of *Montel's theorem*, which yields a criterion for compactness of families of analytic functions.

Compactness is a basic concept in real and complex analysis[1] and an indispensable tool in many existence proofs. For instance, in order to find solutions of an extremal problem, one may start with a set of 'almost extremal' elements. If this set is compact, it contains a converging sequence, and the limit of that sequence is often a good candidate for a solution.

When we wish to apply this *selection principle* to problems involving analytic functions, we need appropriate compactness criteria. For families of *continuous* functions we have the Arcéla–Ascoli theorem (see Rudin [58] p. 145, for instance) which states: a family $\mathcal{F}$ of continuous functions $f : K \to \mathbb{C}$ on a compact set $K \subset \mathbb{C}$ is *relatively compact* (which means that the closure of $\mathcal{F}$ is compact) with respect to uniform convergence on K, if and only if it is *bounded* and *equicontinuous*.

Locally Bounded Families. Somewhat surprisingly, the condition of equicontinuity is superfluous (which is the reason why we do not even define it here) if $\mathcal{F}$ consists of *analytic* functions, and uniform convergence is replaced by *normal convergence* (see Definition 5.1.1). Moreover, even the requirement of boundedness can be relaxed.

Definition 5.2.1. A family $\mathcal{F}$ of functions $f : D \to \mathbb{C}$ on an open set $D \subset \mathbb{C}$ is said to be *locally bounded*, if for any point $a \in D$ there exist a disk $D(a) \subset D$ centered at a and a positive constant $C(a)$ such that

$$|f(z)| \le C(a), \qquad z \in D(a),\ f \in \mathcal{F}. \tag{5.3}$$

Lemma 5.2.2. *Let $\mathcal{F}$ be a locally bounded family of analytic functions on an open set $D \subset \mathbb{C}$, and let K be a compact subset of D. Then $\mathcal{F}$ is* uniformly bounded *and* uniformly Lipschitz continuous *on K, i.e., there exist positive constants C and L such that*

$$|f(z)| \le C, \quad \big|f(z_1) - f(z_2)\big| \le L\,\big|z_1 - z_2\big|, \quad z, z_1, z_2 \in K,\ f \in \mathcal{F}. \tag{5.4}$$

Proof. 1. We cover K by disks $D(a)$ with $a \in K$ according to Definition 5.2.1, select a *finite* covering $D(a_1), \dots, D(a_n)$ of K, and set $C := \max\{C(a_1), \dots, C(a_n)\}$.

2. The union $U := D(a_1) \cup \ldots \cup D(a_n)$ is an open set which contains K, and its boundary has a positive distance to K. Consequently we can decrease the radii of all disks $D(a_k)$ simultaneously by the same small amount δ such that K is still covered by the shrunk disks $\widetilde{D}(a_k)$. Moreover, we can choose δ so that the boundary circles of $\widetilde{D}(a_k)$ are in 'generic position', i.e., there are no pairs of such circles which touch each other and no triples which have a common point.

3. As a result of Step 2 we get an open set $G := \widetilde{D}(a_1) \cup \ldots \cup \widetilde{D}(a_n)$ with $K \subset G \subset \overline{G} \subset U \subset D$. According to the first step, the inclusion $\overline{G} \subset U$ implies that

$$|f(z)| \le C, \qquad z \in \overline{G},\ f \in \mathcal{F}. \tag{5.5}$$

---

[1] Readers who have never heard about compactness are recommended to consult an introductory text on topology.

The boundary J of G consists of a finite number of closed Jordan curves $J_1, \dots, J_m$ which are composed by circular arcs. We choose parameterizations γ_k of J_k which are positively oriented with respect to G (so that G lies to left when we walk along J_k in the direction of increasing parameters t). Then $\Gamma := \gamma_1 + \dots + \gamma_m$ is a cycle which is null-homologous in D (see Definitions 4.3.2 and 4.3.3) and has winding number one about any point in G.

4. Applying the generalized Cauchy integral formula (Theorem 4.3.4) we obtain

$$f(z_0) = \frac{1}{2\pi \mathrm{i}} \int_\Gamma \frac{f(z)}{z - z_0}\, dz, \qquad z_0 \in G,\ f \in \mathcal{F}.$$

In particular we have for arbitrary points $z_1, z_2 \in K$ and $f \in \mathcal{F}$,

$$f(z_1) - f(z_2) = \frac{1}{2\pi \mathrm{i}} \int_\Gamma \frac{f(z)}{z - z_1} - \frac{f(z)}{z - z_2}\, dz = \frac{z_1 - z_2}{2\pi \mathrm{i}} \int_\Gamma \frac{f(z)}{(z - z_1)(z - z_2)}\, dz. \tag{5.6}$$

Since $J = [\Gamma]$ has a positive distance from K and f satisfies the estimate (5.5), the standard integral estimate applied to the integral on the right-hand side shows the existence of a constant L such that

$$\big|f(z_1) - f(z_2)\big| \le L\,\big|z_1 - z_2\big|, \qquad z_1, z_2 \in K,\ f \in \mathcal{F}.$$

□

Pointwise Convergence. It turns out that pointwise convergence and local boundedness of a sequence of analytic functions already imply normal convergence. In fact we prove something stronger.

Theorem 5.2.3. *Assume that (f_n) is a locally bounded sequence of analytic functions $f_n : D \to \mathbb{C}$ in an open set $D \subset \mathbb{C}$. If the sequence (f_n) converges pointwise on a dense subset S of D, then it converges normally in D.*

Proof. Let K be a compact subset of D, and let L denote the Lipschitz constant in the estimate (5.4) for the family $\mathcal{F} := \{f_n\}$ on K. We fix a positive number ε and cover K by a finite collection of open disks $D_1, \dots, D_n$ with radii not exceeding $\varepsilon/(2L)$ such that $D_k \cap K \ne \emptyset$. Since S is dense in D, any disk D_k must contain a point $z_k \in S$. Then it follows that for any $z \in K$ there exists a point z_k such that

$$\big|z - z_k\big| < \varepsilon / L. \tag{5.7}$$

Using the triangle inequality, we estimate $\big|f_n(z) - f_m(z)\big|$ for all $m, n \in \mathbb{N}$ by

$$\big|f_n(z) - f_m(z)\big| \le \big|f_n(z) - f_n(z_k)\big| + \big|f_n(z_k) - f_m(z_k)\big| + \big|f_m(z_k) - f_m(z)\big|$$

with $z_k \in S$ satisfying (5.7). By estimate (5.4) in Lemma 5.2.2, the first and the last term on the right-hand side are less than ε. Since the sequence (f_n) converges for all points in S, the second summand is getting smaller than ε, provided that m and n are sufficiently large, say $m, n \ge N$. In summary we have

$$\big|f_n(z) - f_m(z)\big| < 3\,\varepsilon, \qquad z \in K,\ m, n \ge N. \tag{5.8}$$

By Cauchy's criterion, the sequence (f_n) converges uniformly on K. □

Montel's Theorem. We recall that the limit of a normally convergent sequence of analytic functions is analytic (Theorem 5.1.3). The main result of this section is the following celebrated theorem by Paul Montel.

Theorem 5.2.4 (Montel)**.** *Any locally bounded sequence* (f_n) *of analytic functions* $f_n : D \to \mathbb{C}$ *on an open set* $D \subset \mathbb{C}$ *contains a subsequence which converges normally in* D *to an analytic function* $f : D \to \mathbb{C}$*.*

Proof. 1. The set S of all points in D with rational real and imaginary parts is a dense subset of D. In order to construct a subsequence of (f_n) which converges at all points of S, we utilize a *diagonal argument*.

2. Since S is countable, it can be arranged as a sequence, $S := \{z_1, z_2, \ldots\}$. By assumption, the sequence $\big(f_n(z_1)\big)$ is bounded, so that we can select a subsequence

$$f_{1,1},\; f_{1,2},\; f_{1,3},\; \ldots,\; f_{1,n},\; \ldots, \tag{5.9}$$

which converges at z_1. Analogously, the sequence (5.9) contains a subsequence

$$f_{2,1},\; f_{2,2},\; f_{2,3},\; \ldots,\; f_{2,n},\; \ldots,$$

converging (additionally) at z_2. Continuing in this manner we obtain for any $k \in \mathbb{N}$ a sequence

$$f_{k,1},\; f_{k,2},\; f_{k,3},\; \ldots,\; f_{k,n},\; \ldots, \tag{5.10}$$

which is a subsequence of $(f_{k-1,n})_n$ and which converges at all points $z_1, z_2, \ldots, z_k$.

3. We now form the diagonal sequence,

$$f_{1,1},\; f_{2,2},\; f_{3,3},\; \ldots,\; f_{n,n},\; \ldots. \tag{5.11}$$

Because any 'tail' $f_{k,k},\; f_{k+1,k+1},\; f_{k+2,k+2},\; \ldots$ of that sequence is a subsequence of (5.10), it converges at all points $z_k \in S$. So the diagonal sequence (5.11) satisfies the assumptions of Theorem 5.2.3, and hence it converges normally in D. □

Normal Families. A concise formulation of Montel's theorem uses a standard notion in complex function theory, which gives us another name for relative compactness with respect to normal convergence.

Definition 5.2.5. A family $\mathcal{F}$ of analytic functions on an open set D is said to be *normal*, if any sequence $(f_n) \subset \mathcal{F}$ contains a subsequence which converges normally in D.

Theorem 5.2.6 (Montel)**.** *A family* $\mathcal{F}$ *of analytic functions* $f : D \to \mathbb{C}$ *on an open set* $D \subset \mathbb{C}$ *is normal if and only if* $\mathcal{F}$ *is locally bounded.*

Proof. It only remains to show the "only-if" direction. If $\mathcal{F}$ fails to be locally bounded, there exist a *closed* disk $K \subset D$ and a sequence $(f_n) \subset \mathcal{F}$ such that $\max_{z \in K} |f_n(z)| \geq n$. Because $\mathcal{F}$ is normal, a subsequence (f_{n_k}) converges uniformly on K to a function f which is analytic on D. This leads to a contradiction

$$n_k \leq \max_{z \in K} \big|f_{n_k}(z)\big| \leq \max_{z \in K} \big|f_{n_k}(z) - f(z)\big| + \max_{z \in K} |f(z)| \leq C, \qquad k \in \mathbb{N}.$$

□

5.3 Function Series

By definition, normal convergence of a *function series* means that the sequence of its partial sums converges normally. It follows immediately from Theorem 5.1.3 that normally convergent series of analytic functions can be differentiated term-by-term.

Theorem 4.4.1 tells us that *Laurent series*, and in particular *power series*, are normally convergent in their domains of convergence. We present some other typical examples below.

Fourier Series. Laurent series are in close relation to *Fourier series* of periodic functions. Recall that a function $f : D \to \mathbb{C}$ is periodic with period $\omega \in \mathbb{C}$, if for all $z \in D$ we have $z + \omega \in D$ and $f(z+\omega) = f(z)$.

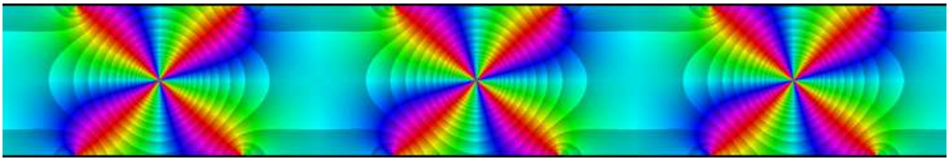


Figure 5.2: A periodic function f in a strip

The first condition is certainly satisfied when the domain set of f is a strip with boundary lines parallel to (the vector) ω. The exponential function

$$w = \exp\left(\frac{2\pi\mathrm{i}\, z}{\omega}\right)$$

maps the strip

$$S := \{z \in \mathbb{C} : a < \operatorname{Im}(z/\omega) < b\} \qquad (5.12)$$

onto the ring domain

$$R := \left\{w \in \mathbb{C} : \mathrm{e}^{-2\pi b} < |w| < \mathrm{e}^{-2\pi a}\right\}.$$

Any function f in the strip S with period ω can be transplanted to a function g in R by

$$g(w) := f\left(\frac{\omega}{2\pi\mathrm{i}} \log w\right),$$

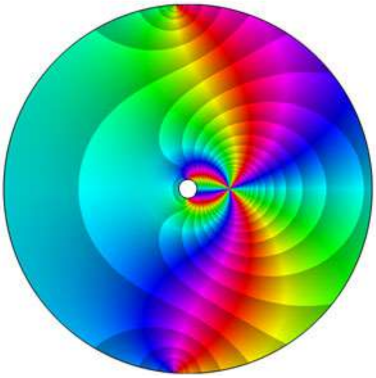


Figure 5.3: Transplantation of the function f to a ring domain

and vice versa, as is demonstrated in Figures 5.2 and 5.3. Due to the periodicity of f it does not matter which branch of the logarithm is chosen, we could even take the principal branch.

If f is analytic in S, then g is analytic in R and can be expanded into a Laurent series

$$g(w) = \sum_{k=-\infty}^{\infty} c_k \, w^k, \qquad w \in R.$$

With the substitution $w = \exp(2\pi \mathrm{i}\, z/\omega)$ we get a corresponding series for the function f,

$$f(z) = \sum_{k=-\infty}^{\infty} c_k \, \mathrm{e}^{2k\pi \mathrm{i}\, z/\omega}, \quad z \in S. \tag{5.13}$$

The series (5.13) is called the (complex) *Fourier series* of the function f. Evaluating the formula (4.56) for the Laurent coefficients with the standard parametrization γ of a circle with radius r and center 0, and translating the result to the variable z by means of the substitution $w = \exp(2\pi \mathrm{i}\, z/\omega)$, we obtain the coefficient formula

$$c_k = \frac{1}{2\pi \mathrm{i}} \int_{\gamma} \frac{g(w)}{w^{k+1}} \, dw = \frac{1}{\omega} \int_{[z_0, z_0+\omega]} f(z) \, \mathrm{e}^{-2k\pi \mathrm{i}\, z/\omega} \, dz, \qquad k \in \mathbb{Z}. \tag{5.14}$$

Here z_0 is an arbitrary point in S and integration is along the segment $[z_0, z_0 + \omega]$. By Cauchy's integral theorem, this segment can be replaced by any path from z_0 to $z_0 + \omega$ in S. We summarize the results in the following theorem.

Theorem 5.3.1. *If $f : S \to \mathbb{C}$ is analytic and ω-periodic in the strip S defined by (5.12), then it can be represented by the Fourier series (5.13) with coefficients given by (5.14). The Fourier series converges absolutely and normally in S.*

If $\omega = 2\pi$ and S contains the real axis, we obtain (formally) the classical Fourier series of 2π-periodic functions on the real line.

Example 5.3.1 (Jacobi Theta Functions). Let q be a complex number with $|q| < 1$. Then the Laurent series

$$f_q(z) = \sum_{k=-\infty}^{\infty} q^{k^2} \, z^k, \tag{5.15}$$

converges in the *punctured plane* $\dot{\mathbb{C}} := \mathbb{C} \setminus \{0\}$. The substitution $w = \exp(2\pi \mathrm{i}\, z)$ transforms the functions f_q to 1-periodic functions $\vartheta_q(z) := f_q\big(\exp(2\pi \mathrm{i}\, z)\big)$ in the plane. The function ϑ_q is the *Jacobi Theta Function* with *nome* q which we already met in Example 3.3.3.

Figure 5.4 depicts the phase portraits of these functions with nome $q = 0.9$ and $q = 0.9\, \mathrm{e}^{\mathrm{i}\pi/6}$ in the square $|\mathrm{Re}\, z| < 0.6$, $|\mathrm{Im}\, z| < 0.6$.

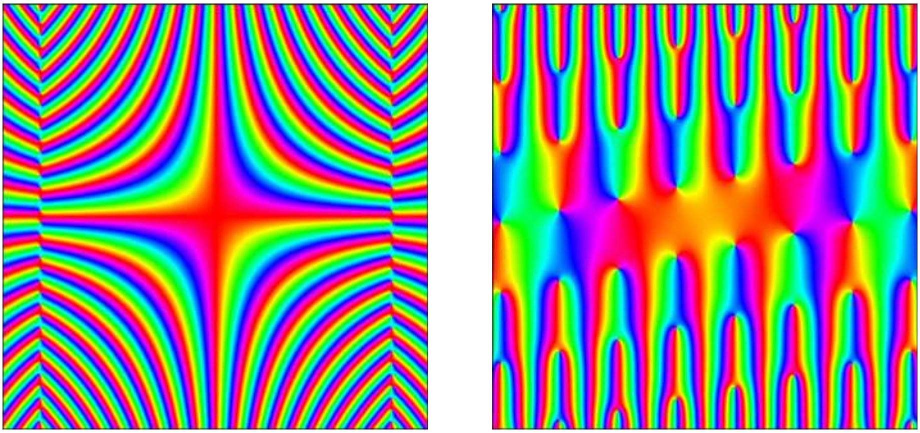


Figure 5.4: Two Jacobi Theta Functions ϑ_q with $q = 0.9$ and $q = 0.9\,\mathrm{e}^{\mathrm{i}\pi/6}$

Lambert Series. Another interesting class of function series is constituted by the so-called *Lambert series*, which have the general form

$$\sum_{k=1}^{\infty} a_k \, \frac{z^k}{1 - z^k}. \tag{5.16}$$

If infinitely many coefficients are different from zero (and this is the only case of interest) the poles of the partial sums of this series include a set of roots of unity which are dense on the unit circle, so that the unit circle forms a natural boundary. Convergence of Lambert series is therefore usually considered only in the unit disk, though the series may converge for other values of z as well.

Theorem 5.3.2. *Let* $\alpha := \limsup \sqrt[k]{|a_k|}$. *Then the Lambert series* (5.16) *converges absolutely and normally in the disk*

$$\{z \in \mathbb{C} : |z| < R := \min(1, 1/\alpha)\}. \tag{5.17}$$

Proof. Let $0 < r < R \le 1$. The definition of R guarantees that there exist a positive number q and an integer K such that

$$r \sqrt[k]{|a_k|} \le q < 1, \qquad k \ge K.$$

Consequently, the summands of (5.16) can be uniformly estimated for all z with $|z| \le r < R$,

$$\left| a_k \, \frac{z^k}{1 - z^k} \right| \le |a_k| \, \frac{|r|^k}{1 - r} \le \frac{|q|^k}{1 - r},$$

so that the series converges absolutely and uniformly by the root criterion. □

In particular the Lambert series defines an analytic function in the unit disk if the sequence of coefficients a_k is bounded.

Example 5.3.2. Lambert series are important in number theory. A famous example is the series

$$f(z) := \sum_{k=1}^{\infty} \frac{z^k}{1-z^k}, \tag{5.18}$$

which converges normally in the unit disk. The application we have in mind is related to the power series expansion of f at the origin. To find this series, we mention that $1/(1-z^k)$ is the sum of a geometric series with quotient z^k, and observe that the term

$$\frac{z^k}{1-z^k} = z^k + z^{2k} + z^{3k} + \dots \tag{5.19}$$

contributes one unit to the Taylor coefficient a_n of f if and only if k is a divisor of n. Summing up over all k, we get that a_n is equal to the number $d(n)$ of positive integral divisors of n. Note that the order of summation of the series (5.18) and (5.19) can be interchanged since the series converge absolutely.

In number theory, the counting function d is called the *divisor function*. Saying that f is the *generating function* of the divisor function, just means that the Taylor coefficients a_n of f coincide with the values $d(n)$ of the divisor function for all n.

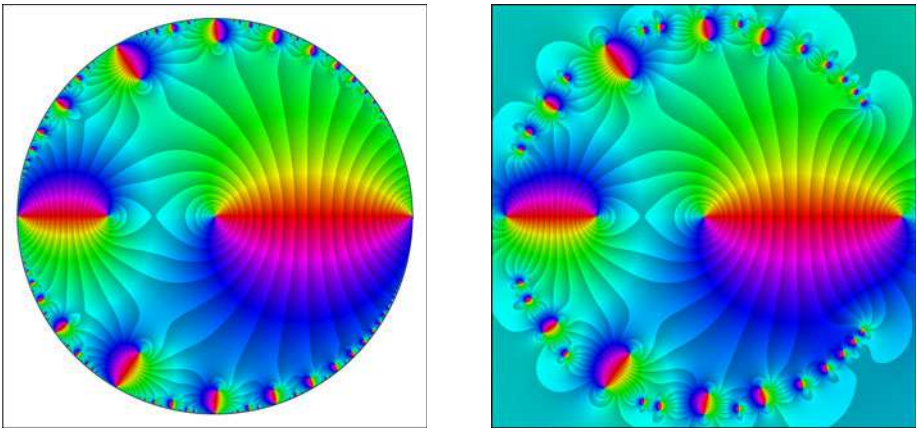


Figure 5.5: Phase portraits of f defined in (5.18) and its partial sum s_{10}

Figure 5.5 (right) shows the enhanced phase portrait of the partial sum s_{10} of (5.18), the function f is depicted on the left. Note that s_n converges (normally) to infinity in the exterior of the unit disk $|z| > 1$.

Dirichlet Series. The last general class of series which we are now going to consider is formed by the so-called *Dirichlet series*,

$$f(z) := \sum_{n=1}^{\infty} \frac{a_n}{n^z} = \sum_{n=1}^{\infty} a_n \exp(-z \operatorname{Log} n). \tag{5.20}$$

Since the absolute value of $\exp(z \operatorname{Log} n)$ is determined by the real part of z, we expect that the convergence of Dirichlet series is mainly dependent on $\operatorname{Re} z$. This is indeed the case: if (5.20) converges for some z_0, then it converges for all z with $\operatorname{Re} z > \sigma := \operatorname{Re} z_0$. Convergence is uniform in any sector $|\arg(z - z_0)| \leq a < \pi/2$, but in general *not* in the half plane $\operatorname{Re} z > \operatorname{Re} z_0$. For a proof of these facts and the following theorem we refer to Titchmarsh [65], Chapter 9.

Theorem 5.3.3. *Let (a_k) be a sequence of complex numbers. Then there exists a number $\sigma_c \in \mathbb{R} \cup \{-\infty, +\infty\}$ such that the Dirichlet series (5.20) converges for all z with $\operatorname{Re} z > \sigma_c$ and diverges for all z with $\operatorname{Re} z < \sigma_c$. The* abscissa of convergence *σ_c is given by*

$$\sigma_c = \begin{cases} \limsup\limits_{n\to\infty} \dfrac{\operatorname{Log} |a_1 + a_2 + \cdots + a_n|}{\operatorname{Log} n} & \text{if } \sum a_k \text{ is divergent,} \\ \limsup\limits_{n\to\infty} \dfrac{\operatorname{Log} |a_{n+1} + a_{n+2} + \cdots|}{\operatorname{Log} n} & \text{if } \sum a_k \text{ is convergent.} \end{cases} \tag{5.21}$$

Moreover, the series converges normally in the half-plane $\operatorname{Re} z > \sigma_c$.

One might expect that, in analogy to power series, Dirichlet series also converge absolutely in their half-plane of convergence. This is not so in general, there may be a strip where the series converges conditionally. More precisely, there is an *abscissa of absolute convergence* $\sigma_a \geq \sigma_c$, such that the series converges absolutely if $\operatorname{Re} z > \sigma_a$, and does not converge absolutely if $\operatorname{Re} z < \sigma_a$. The value of σ_a can be determined from (5.21) with a_k replaced by $|a_k|$.

Example 5.3.3 (The Riemann Zeta Function). The most famous Dirichlet series defines the celebrated *Riemann Zeta function*,

$$\zeta(z) := \sum_{n=1}^{\infty} \frac{1}{n^z}, \qquad \operatorname{Re} z > 1. \tag{5.22}$$

Since $|n^z| = n^{\operatorname{Re} z}$, so that $|n^{-z}| \leq n^{-\sigma}$ for all z with $\operatorname{Re} z \geq \sigma > 1$, the series (5.22) converges (absolutely) in the half plane $\operatorname{Re} z > 1$ (this also follows immediately from Theorem 5.3.3). For $z = 1$ the series represents the divergent harmonic series. Another remarkable fact is

$$\zeta(2) = 1 + \frac{1}{2^2} + \frac{1}{3^2} + \ldots + \frac{1}{n^2} + \ldots = \frac{\pi^2}{6}.$$

The picture on the left in Figure 5.6 shows a phase portrait of the partial sum s_{1000} of the series (5.22) in the square $|\operatorname{Re} z| < 10$, $|\operatorname{Im} z| < 10$. In the left part of this square, where the series does not converge, the modulus of s_{1000} is quite large, the partial sums are growing exponentially in the direction of the negative real axis (which is related to the parallel strips in the phase portrait, see comments on page 148).

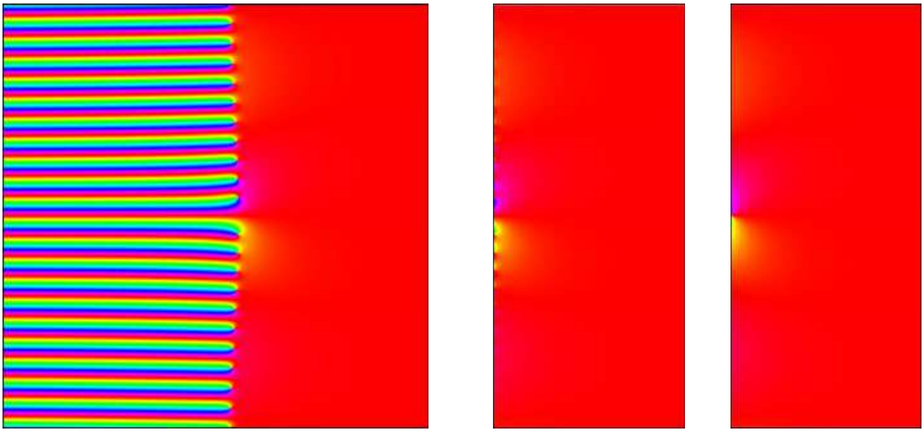


Figure 5.6: Divergence and convergence of the Dirichlet series for the Zeta function

The window in the middle shows the same partial sum in the subdomain where the series converges. The picture on the right is the corresponding phase portrait of the sum of the series. It is clearly visible that convergence is rather slow if $\operatorname{Re} z$ is close to one, even s_{1000} is still a very poor approximation of the Zeta function. In particular, the few details which show up in the middle of the left boundary are misleading, and have (almost) no relation to the actual function.

Though the Zeta function looks quite uninteresting in this domain, it is of an unbelievable richness and conceals a number of deep secrets, which sometimes border on the mysterious. Among them is one of the monumental open problems of mathematics: the *Riemann hypothesis* on the location of its zeros.

We shall continue the investigation of the Riemann Zeta function in Example 5.6.3 of Section 5.6, where we show that it has an analytic extension to the complex plane, with a simple pole at $z_0 = 1$.

Doubly Infinite Series. As was the case for Laurent series, the (absolute, uniform, normal, ...) convergence of general *doubly infinite series* is reduced to the convergence of two one-sided series by the splitting

$$\sum_{k=-\infty}^{\infty} f_k := \sum_{k=1}^{\infty} f_{-k} + \sum_{k=0}^{\infty} f_k. \tag{5.23}$$

Example 5.3.4 (The Cotangent Function Revisited). We show that the doubly infinite series, with a separately written summand corresponding to $k = 0$,

$$f(z) := \frac{1}{z} + \sum_{k\in\mathbb{Z}\setminus\{0\}} \left(\frac{1}{z-k} + \frac{1}{k} \right) \tag{5.24}$$

converges absolutely and normally in $\mathbb{C}\setminus\mathbb{Z}$. If K is any compact subset K of $\mathbb{C}\setminus\mathbb{Z}$ there exists an R such that $|z| \le R$ for all $z \in K$. Then

$$\left| \frac{1}{z-k} + \frac{1}{k} \right| = \left| \frac{z}{k(z-k)} \right| \le \frac{2\,R}{k^2}, \tag{5.25}$$

for all k with $|k| \ge 2\,R$, so that $2R \sum 1/k^2$ is a convergent majorant series for (5.24).

Figure 5.7 depicts phase portraits of partial sums $s_m^n := \sum_{k=m}^n f_k$ of the doubly infinite series (5.24) in the square $|\mathrm{Re}\, z| < 3$, $|\mathrm{Im}\, z| < 3$. On the left picture we see s_0^{20}, the picture on the right shows s_{-20}^{20}.

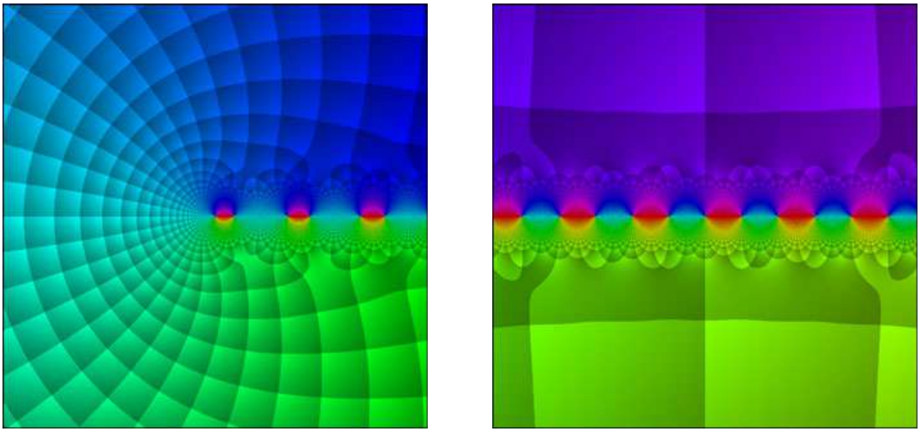


Figure 5.7: Partial sums s_0^{20} and s_{-20}^{20} of the doubly infinite series (5.24)

Since the (double) series converges absolutely, it can be rearranged by collecting the summands corresponding to $k = n$ and $k = -n$,

$$f(z) = \frac{1}{z} + \sum_{k=1}^{\infty} \frac{2z}{z^2 - k^2}. \tag{5.26}$$

Attentive readers may have observed that the illustration on the right in Figure 5.7 resembles the phase portrait of the cotangent function in Figure 3.20, which we reproduce for convenience again in Figure 5.8. And indeed the sum of the series (5.26) *is* a scaled cotangent function, as we shall show a little later.

Another interesting observation can be made by looking at the phase portrait of the partial sums from afar. Figure 5.8 (right) shows such a view for s_{-20}^{20} of (5.24) in a square with edge length 60.

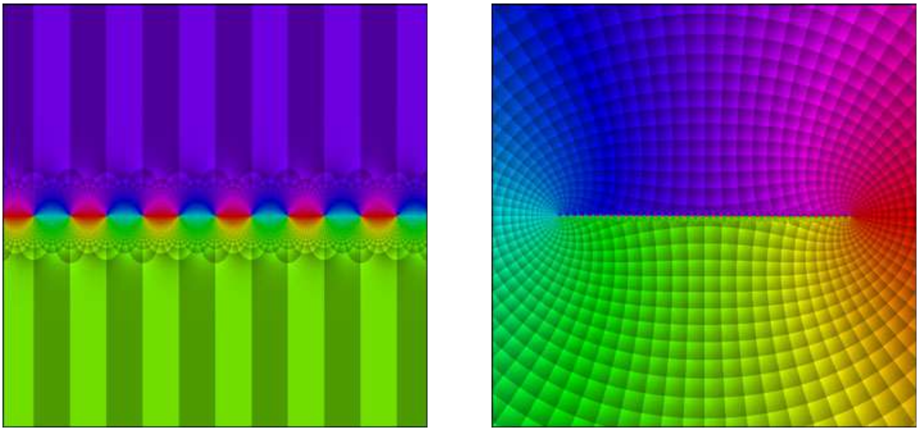


Figure 5.8: The cotangent function and a view from afar of a partial sum of (5.26)

We shall see a quite similar phase portrait in Figure 5.12 of Section 5.5 when we study Cauchy integrals with constant density on an interval. This is of course not accidental – up to some constant factor and an appropriate scaling of the plane, the partial sums s_{-n}^{n} are just the approximations of that Cauchy integral by the rectangle rule.

Now let us confirm that the function f defined by (5.26) is related to the cotangent function. First of all we need a more precise conjecture about what this relation could be. A useful tool to achieve this goal is the principle of *pole matching*. The cotangent function has (simple) poles at the points $z = k\pi$, $k \in \mathbb{Z}$, while f has (simple) poles at the integers. To match the poles we thus compare f with $\cot(\pi z)$. Indeed, the phase portraits of these two functions seem to be identical. Taking this for granted, it follows that $f(z) = c\,\cot(\pi z)$ with $c \in \mathbb{R}_+$. The positive constant c cannot be estimated from the phase portrait. Instead we consider a Laurent expansion of f and $c\cot(\pi z)$ at the origin. Since $f(z) - 1/z$ is analytic at $z_0 = 0$ we have $f(z) = 1/z + a_0 + a_1 z + \ldots$. From $\sin z = z\,(1 - z^2/6 \pm \ldots)$ and $\cos z = 1 - z^2/2 \pm \ldots$ we deduce that $\cot z = 1/z + b_0 + b_1 z + \ldots$, which implies $c\,\cot(\pi z) = (c/\pi)1/z + c_0 + \ldots$. Comparing coefficients, we find $c = \pi$, so that our qualified working hypothesis is

$$\pi\,\cot(\pi z) = \frac{1}{z} + \sum_{k=1}^{\infty} \frac{2z}{z^2 - k^2}. \tag{5.27}$$

In order to prove this conjecture we set $g(z) := \pi\,\cot(\pi z)$ and consider the function $h := f - g$. If follows as in the preceding paragraph for $z_0 = 0$ that *all* poles $z_k = k$

of f and g compensate each other, so that h is an entire function.

In the next step we are going to show that h is bounded. A direct estimation of $|h|$ would be rather technical, so let us first show that the *derivative* h' is bounded on $\mathbb{C}$. This estimate of $h'(z)$ depends on the location of z. If z is close to the real axis we can avoid any computation by a clever application of the maximum principle in the (closed) rectangle $R := \{x + \mathrm{i}\, y : |x| \leq 1/2,\ |y| \leq 1\}$.

Because f' and g' are analytic in $\mathbb{C} \setminus \mathbb{Z}$, they are continuous on the (compact) boundary ∂R of R. By Weierstrass' theorem, f', g', and hence h' are bounded on ∂R, and the maximum principle implies that h' is bounded on R. Since h' is 1-periodic, it is bounded on the strip $|\mathrm{Im}\, z| \leq 1$.

In order to estimate $h'(z)$ for $|\mathrm{Im}\, z| > 1$ we differentiate f and g. Normal convergence on $\mathbb{C} \setminus \mathbb{Z}$ guarantees that (5.24) can be differentiated term-by-term, which yields the representation

$$h'(z) = f'(z) - g'(z) = \frac{\pi^2}{(\sin \pi z)^2} - \sum_{k=-\infty}^{\infty} \frac{1}{(z-k)^2}, \qquad z \in \mathbb{C} \setminus \mathbb{Z}. \tag{5.28}$$

It is remarkable, that the convergence of the series (5.24) is *improved by differentiation*, which is the reason why we consider h' and not h itself.

Writing $z = x + \mathrm{i}\, y$, and assuming $|x| \leq 1/2$, $|y| \geq 1$ we estimate

$$|z-k|^2 = k^2 - 2xk + x^2 + y^2 \geq (|k| - 1/2)^2 - 1/4 + x^2 + y^2 \geq (|k| - 1/2)^2 + 3/4. \tag{5.29}$$

Consequently for $|x| \leq 1/2$, and then by periodicity for all $x \in \mathbb{R}$,

$$|f'(x + \mathrm{i}\, y)| \leq \sum_{k=-\infty}^{\infty} \frac{1}{(|k| - 1/2)^2 + 3/4} \leq C < \infty, \qquad |y| \geq 1. \tag{5.30}$$

The estimate of g' is straightforward,

$$|\sin(\pi z)|^2 = \frac{1}{4}\left(\mathrm{e}^{2\pi y} + \mathrm{e}^{-2\pi y}\right) - \frac{1}{2}\cos(2\pi x) \geq \frac{1}{4}\left(\mathrm{e}^{2\pi |y|} - 2\right). \tag{5.31}$$

Consequently,

$$|g'(x + \mathrm{i} y)| \leq \frac{4\pi^2}{\mathrm{e}^{2\pi|y|} - 2} \leq \frac{1}{10}, \qquad |y| \geq 1. \tag{5.32}$$

The estimates (5.30) and (5.32) show that h' is also bounded for $|\mathrm{Im}\, z| > 1$, and thus on the entire plane. Using Liouville's theorem we conclude that $h' = \mathrm{const}$. Letting $y \to \infty$ in (5.30) and (5.32) we get $h' = 0$, and hence

$$\frac{\pi^2}{(\sin \pi z)^2} = \sum_{k=-\infty}^{\infty} \frac{1}{(z-k)^2}. \tag{5.33}$$

Once we know that h is constant, $h = 0$ follows because f, g, and hence h are odd functions.

Figure 5.9 visualizes the function $\pi^2/(\sin \pi z)^2$ and the partial sum s_{-100}^{100} of the series (5.33). The convergence of this series is rather slow, with an error of order $1/n$ for the nth partial sum. The reason for the discrepancy in the phase portraits is the fast decay of the function with increasing distance from the real axis (which can be seen more clearly in the domain coloring picture on the right-hand side), which causes large deviation in the phase even when the difference of the function values is quite small.

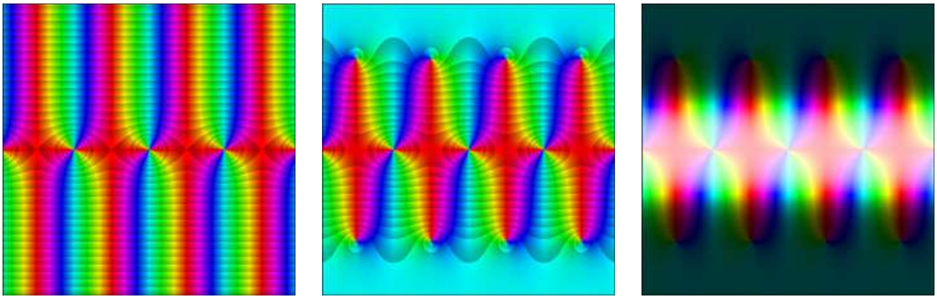


Figure 5.9: The function $\pi^2/(\sin \pi z)^2$ and the partial sum s_{-100}^{100} of (5.33)

The representations (5.27) and (5.33) are prototypes of *partial fraction decompositions* of analytic functions, which we shall study in more detail in Volume 2.

5.4 Infinite Products

When one is interested in constructing analytic functions with prescribed behavior, some goals can be more easily achieved using products instead of sums. In this section we therefore turn our attention towards *infinite products* of complex numbers and functions.

Convergence of Infinite Products. Recalling the definition of a convergent series, it is tempting to say that an infinite product

$$\prod_{k=1}^{\infty} c_k \tag{5.34}$$

converges, if the sequence (p_n) of its *partial products* $p_n := \prod_{k=1}^{n} c_k$ converges. However, closer inspection of this setting shows that it would have several unpleasant consequences. For example, if one single factor c_k is zero, the product would converge, regardless of what the other factors are. This is also contrary to our expectation that convergence or divergence of a product (or a series) only depends on its "tail" and not on the initial terms. Another property of finite products which we would like to preserve is that a product is different from zero if all its factors are non-zero. The above definition would violate this rule, for example if $c_k = 1/k$. These observations motivate the following more subtle definition.

Definition 5.4.1. Let (c_k) be a sequence of complex numbers. We say that the infinite product (5.34) *converges*, if at most finitely many factors c_k are zero (that is $c_k \neq 0$ for all $k > K$), the limit

$$p(K) := \lim_{n\to\infty} \prod_{k=K+1}^{n} c_k \tag{5.35}$$

exists, and is finite and non-zero. If the infinite product converges, its *value* (or, well, its *product*) p is defined by

$$p := p_1 \cdot p_2 \cdot \ldots \cdot p_K \cdot p(K). \tag{5.36}$$

A product (5.34) which does not converge is said to be *divergent*.

It is clear that the value of a convergent product does not depend on the choice of the integer K.

Like for series, the symbol $\prod c_k$ also denotes the value of a convergent product. It follows almost immediately from the definition that convergent products "can be multiplied" according to the rule

$$\left(\prod_{k=1}^{\infty} c_k\right)\left(\prod_{k=1}^{\infty} d_k\right) = \prod_{k=1}^{\infty}(c_k d_k).$$

For $k > K$ we write the factor c_k as $p(k-1)/p(k)$ and obtain that $c_k \to 1$ is a *necessary condition* for convergence of the infinite product (5.34). This observation forms the background for writing infinite products in the form

$$\prod_{k=1}^{\infty}(1 + a_k). \tag{5.37}$$

A Convergence Criterion. A necessary and *sufficient condition* for convergence of infinite products is formulated in the next theorem, which reduces the problem to the convergence of an associated infinite series. Since, according to Definition 5.4.1, the leading factors $c_1, \ldots, c_K$ are irrelevant for the convergence of the product, we may restrict ourselves to the case where all c_k are different from zero. Recall that Log denotes the principal branch of the complex logarithm.

Theorem 5.4.2. *Let (c_k) be a sequence of non-zero complex numbers. Then the infinite product and the infinite sum*

$$\prod_{k=1}^{\infty} c_k, \qquad \sum_{k=1}^{\infty} \operatorname{Log} c_k \tag{5.38}$$

either both converge or both diverge. If s is the sum of the series, then the product has the value $\exp s$.

Proof. 1. Let $a_k := \operatorname{Log} c_k$. We consider the partial products and partial sums

$$p_n := \prod_{k=1}^{n} c_k, \qquad s_n := \sum_{k=1}^{n} a_k.$$

Since $c_k = \exp a_k$ we have $p_n = \exp s_n$, which shows that the convergence of the sum (to s) implies convergence of the product (to $\exp s$).

2. Now assume that the product converges to $p \neq 0$. As a consequence of $p_n = \exp s_n$ we then obtain that $\operatorname{Re} s_n = \operatorname{Log} |p_n| \to \operatorname{Log} p$. Showing that the imaginary part of s_n also converges is a little more delicate, since $\operatorname{Im} s_n$ is determined by $\exp s_n$ only up to an additive value of $2m\pi$.

But there is some additional information: since the product converges, c_k must converge to 1, so that a_k converges to zero. Then $\operatorname{Im} s_n - \operatorname{Im} s_{n-1} = \operatorname{Im} a_n$ tends to zero as $n \to \infty$, so that

$$|\operatorname{Im} s_n - \operatorname{Im} s_{n-1}| < \pi/2, \qquad n \geq N, \tag{5.39}$$

for some positive integer N. Since the points p_n converge to $p \neq 0$, there is some sector S with opening angle $\pi/2$ and vertex 0 such that p_n lies in S if n is sufficiently large. Increasing N, if necessary, we may assume that this happens for all $n \geq N$. More precisely, for every $n \geq N$ either

$$|\operatorname{Im} s_n - \operatorname{Im} s_N| < \pi/2 \quad \text{or} \quad |\operatorname{Im} s_n - \operatorname{Im} s_N| > 3\pi/2. \tag{5.40}$$

In fact the first inequality of (5.40) must hold for all $n \geq N$. Otherwise there is a smallest value n for which the second inequality would be satisfied. Then, by the triangle inequality,

$$|\operatorname{Im} s_n - \operatorname{Im} s_{n-1}| > |\operatorname{Im} s_n - \operatorname{Im} s_N| - |\operatorname{Im} s_{n-1} - \operatorname{Im} s_N| > \pi,$$

in contradiction to (5.39). Finally, the conditions

$$|\operatorname{Im} s_n - \operatorname{Im} s_N| < \pi/2, \qquad \exp s_n \to p$$

ensure the convergence of $\operatorname{Im} s_n$. ☐

Absolute Convergence. To allow certain manipulations with infinite products, like rearrangement of the factors, the concept of convergence must be strengthened. The analogue of absolute convergence for sums is introduced in the next definition.

Definition 5.4.3. The infinite product $\prod_{k=1}^{\infty}(1 + a_k)$ is *absolutely convergent* if $\prod_{k=1}^{\infty}(1 + |a_k|)$ is convergent.

The following necessary and sufficient criterion transforms the problem of testing absolute convergence of products to checking convergence of the associated series. As in Theorem 5.4.2 we confine ourselves to the case where all factors are non-zero.

Theorem 5.4.4. *Let* (a_k) *be a sequence of complex numbers with* $a_k \neq -1$. *Then the infinite product* $\prod_{k=1}^{\infty}(1+a_k)$ *converges absolutely if and only if one (and then all) of the following three series converges,*

$$\sum_{k=1}^{\infty} |\operatorname{Log}(1+a_k)|, \qquad \sum_{k=1}^{\infty} \operatorname{Log}(1+|a_k|), \qquad \sum_{k=1}^{\infty} |a_k|. \tag{5.41}$$

Proof. The result can be easily derived from Theorem 5.4.2, the comparison test, and the estimate

$$|z|/2 \leq |\operatorname{Log}(1+z)| \leq 2\,|z|, \qquad |z| \leq 1/2. \tag{5.42}$$

$\square$

By virtue of Theorem 5.4.2 and the criterion of Theorem 5.4.4 with the first sum in (5.41), we conclude that any absolutely convergent product is convergent.

To convince ourselves that the factors of an absolutely convergent product can be rearranged in any order without changing the value p of the product, we refer to the representation $p = \exp s$ in Theorem 5.4.2. Since now s is the sum of an *absolutely* convergent series, it is invariant with respect to permutations of the summands, so that also p does not depend on the order of the factors.

Products of Functions. Now let (f_k) be a sequence of complex functions on a domain set D. We say that the infinite product $\prod_{k=1}^{\infty} f_k$ *converges pointwise* (*absolutely*) on a set $A \subset D$, if $\prod_{k=1}^{\infty} f_k(z)$ converges (absolutely) for each $z \in A$. In this case

$$f(z) := \prod_{k=1}^{\infty} f_k(z), \qquad z \in A, \tag{5.43}$$

defines a function $f : A \to \mathbb{C}$. Note that the definition of infinite products guarantees that the zero set of f is just the union of the zero sets of the factors f_k.

Uniform and Normal Convergence. Pointwise convergence is a rather weak form of convergence, and for many purposes it has to be strengthened. Again, *normal convergence* is the most adequate concept which satisfies our needs.

Definition 5.4.5. The infinite product (5.43) is said to *converge uniformly* on a set A if the following three conditions are satisfied:

(i) the product converges pointwise on A,
(ii) there exists an integer K such that f_k has no zero in A for all $k > K$,
(iii) the sequence (p_n) of truncated partial products

$$p_n(z) := \prod_{k=K+1}^{n} f_k(z) \tag{5.44}$$

converges uniformly on A.

We declare that (5.43) *converges normally* in D, if it converges uniformly on any compact subset of D.

We emphasize that this definition of normal convergence guarantees that the value of a normally convergent infinite product of analytic functions is analytic. The most important criterion for uniform convergence of infinite products is the following result, which translates the problem into the language of series.

Theorem 5.4.6. *Assume that every function $f_1, f_2, \ldots$ is bounded on A, and suppose that the series*

$$\sum_{k=1}^{\infty} |f_k - 1| \tag{5.45}$$

converges uniformly in A. Then the product $\prod\limits_{k=1}^{\infty} f_k$ converges absolutely and uniformly in A.

Proof. 1. The uniform convergence of the sum (5.45) implies that the functions f_k have no zeros for sufficiently large k. Hence, by Theorem 5.4.4, the product $\prod f_k$ converges pointwise (absolutely) on A to a function p.

2. Let $g_k := f_k - 1$. Uniform convergence of $\sum |g_k|$ implies the existence of an integer K such that

$$\sum_{k=K+1}^{\infty} |g_k(z)| < 1/2, \qquad z \in A. \tag{5.46}$$

In particular, $|g_k| < 1/2$, so that $f_k = 1 + g_k$ has no zeros in A if $k > K$. Furthermore (see (5.42)),

$$|\operatorname{Log} f_k(z)| = |\operatorname{Log}(1 + g_k(z))| \le 2\,|g_k(z)|, \quad z \in A,\ k > K. \tag{5.47}$$

3. Representing the truncated partial products p_n defined in (5.44) by

$$p_n(z) = \exp\Bigg(\sum_{k=K+1}^{n} \operatorname{Log} f_k(z)\Bigg), \tag{5.48}$$

and using (5.46) and (5.47), we obtain the estimate

$$|p_n(z)| = \exp\Bigg(\sum_{k=K+1}^{n} \operatorname{Re}\operatorname{Log} f_k(z)\Bigg) \le \exp\Bigg(2 \sum_{k=K+1}^{n} |g_k(z)|\Bigg) \le \mathrm{e}, \tag{5.49}$$

which is valid for all $z \in A$. Taking the limit $n \to \infty$ it follows that the truncated products p_n converge to a *bounded* function. Since the factors $f_1, \ldots, f_K$ are *assumed* to be bounded, the value p of the product is a bounded function.

4. Given any positive number $\varepsilon < 1$, uniform convergence of $\sum |g_k|$ ensures that we can increase K, if necessary, such that $\sum_{k=m+1}^{n} |g_k(z)| < \varepsilon$ for all $n > m > K$

and $z \in A$. Consequently,

$$\left| \sum_{k=m+1}^{n} \operatorname{Log} f_k \right| \leq 2 \sum_{k=m+1}^{n} |g_k| < 2\varepsilon,$$

which shows that $p_m(z)/p_n(z) = \exp\bigl(r_{mn}(z)\bigr)$ with $|r_{mn}(z)| < 2\varepsilon$. When n tends to infinity we get

$$\left| \frac{p_m(z)}{p(z)} - 1 \right| < 4\varepsilon, \qquad z \in A,$$

where we made use of the estimate $|\exp z - 1| < 2\,|z|$ for $|z| < 1$. This shows that p_m/p converges uniformly to 1 as $m \to \infty$, and since p is bounded, p_m converges uniformly to p, as claimed. ☐

For applications in complex function theory, it is important to know under which conditions the analyticity of an infinite product follows from the analyticity of its factors. One such criterion is provided by the following corollary of Theorem 5.4.6.

Corollary 5.4.7. *If the functions f_k are analytic in D and $\sum |f_k - 1|$ converges normally in D, then the product $\prod f_k$ converges normally (and absolutely) to an analytic function in D.*

Proof. If A is a compact subset of D, then every analytic function f_k is automatically bounded on A, so that Theorem 5.4.6 applies to the restriction of f_k to A. Since A can be chosen arbitrarily, the assertion follows. ☐

Infinite Blaschke Products. We continue the investigation of Blaschke products in Example 3.4.4, now admitting that they are composed of an infinite number of Blaschke factors

$$f_k(z) = c_k \,\frac{z - z_k}{1 - \overline{z}_k\, z}, \qquad |c_k| = 1,\ 0 < |z_k| < 1. \tag{5.50}$$

The crucial problem is to find criteria which ensure that a Blaschke product with infinitely many factors f_k converges. To begin with, we restrict all functions to the unit disk and forget about the exterior of the unit circle where the Blaschke factors have poles.

First of all we consider the necessary condition that the f_k must tend to one. To bring at least $f_k(0)$ as close as possible to 1, we *normalize* the Blaschke factors (5.50) such that $f_k(0)$ is positive by choosing an appropriate unimodular constant c_k,

$$f_k(z) = \frac{|z_k|}{z_k}\,\frac{z_k - z}{1 - \overline{z}_k\, z}, \qquad 0 < |z_k| < 1. \tag{5.51}$$

Figure 5.10 shows the effect of normalization for three Blaschke factors with zeros at -0.7, $0.9\,\mathrm{i}$, and 0.96.

Evaluating the product $\prod f_k(z)$ at $z = 0$ we obtain $\prod |z_k|$, which by Theorem 5.4.4 converges if and only if the z_k satisfy the condition

$$\sum_{k=1}^{\infty} (1 - |z_k|) < +\infty. \tag{5.52}$$

This so-called *Blaschke condition* tells us how fast the zeros of the Blaschke factors must tend to the boundary of the unit disk in order to make the product converge, at least at $z = 0$.

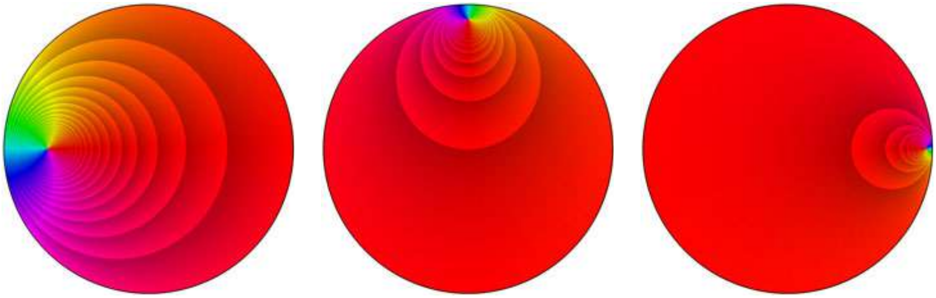


Figure 5.10: Phase portraits with contour lines of normalized Blaschke factors

Now let us check how successful we have been for the other values of z:

$$\left| 1 - \frac{|z_k|}{z_k} \frac{z_k - z}{1 - \overline{z}_k\, z} \right| = (1 - |z_k|) \left| \frac{1 + z\, |z_k|/z_k}{1 - \overline{z}_k\, z} \right| \le \frac{1 + |z|}{1 - |z|} \big(1 - |z_k|\big). \tag{5.53}$$

Together with Theorem 5.4.6 this estimate shows that the Blaschke condition (5.52) is also *sufficient* for *absolute* and *normal* convergence of a Blaschke product in the unit disk.

If we finally add some Blaschke factors with a zero at the origin, which have been excluded so far since they cannot be normalized, we can summarize our findings in the following theorem.

Theorem 5.4.8. *Let (z_k) be an infinite sequence of complex numbers with $|z_k| < 1$, which is arranged such that $z_1 = \ldots = z_m = 0$ (with $m \ge 0$) and $z_k \ne 0$ for $k > m$. Then the Blaschke condition (5.52) is necessary and sufficient for the convergence of the Blaschke product*

$$z^m \prod_{k=m+1}^{\infty} \frac{|z_k|}{z_k} \frac{z_k - z}{1 - \overline{z}_k\, z}, \tag{5.54}$$

at $z = 0$. If (5.52) is satisfied, then the Blaschke product (5.54) converges absolutely and normally in $\mathbb{D}$ to a holomorphic function f whose zeros are exactly the points z_k (counted according to multiplicity). Moreover, the modulus of f is bounded by one, $|f(z)| < 1$ for all $z \in \mathbb{D}$.

Proof. The Blaschke condition and the estimate (5.53) guarantee that the series $\sum |f_k - 1|$ converges *uniformly* on any disk $|z| \le r < 1$, and hence on any compact subset of $\mathbb{D}$. The first assertion then follows from Corollary 5.4.7. By definition of infinite products, the zero set of f is the union of the zero sets of the (Blaschke) factors. Finally, $|f(z)| < 1$ follows from $|f_k(z)| \le 1$ and $|f_1(z)| < 1$ on $\mathbb{D}$. ☐

In Figure 5.11 we have tried to visualize two infinite Blaschke products. Of course, such pictures cannot show infinitely many zeros, but the special patterns help to acquire a feeling how these functions behave close to the boundary.

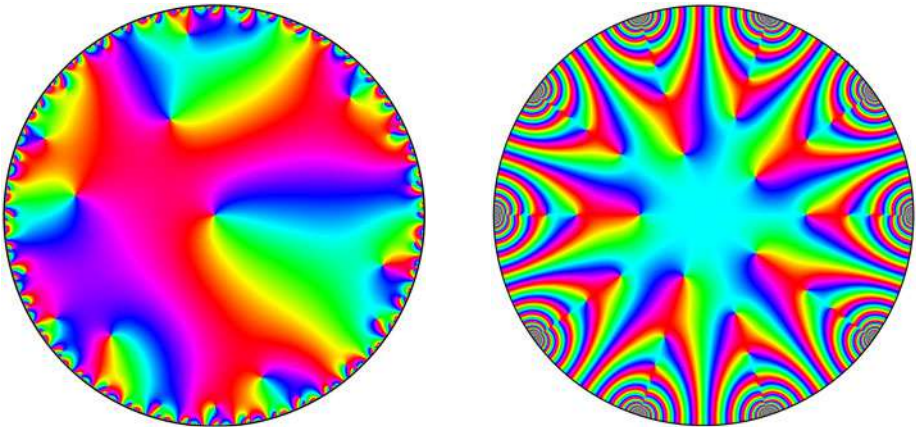


Figure 5.11: Phase portraits of two infinite Blaschke products

The function depicted on the left has zeros $z_k = \big(1 - k^{-1.01}\big)\, \mathrm{e}^{\mathrm{i}\, k}$ with $k = 1, 2, \ldots$, while for the function on the right-hand side

$$z_{kj} = (-1)^k\, \omega^j \left(1 - \frac{25}{(25+k)^2}\right), \qquad \omega = \mathrm{e}^{2\pi \mathrm{i}/5}, \quad j = 0, \ldots, 4, \quad k = 1, 2, \ldots .$$

Note that the second function is not normalized as usual but has an additional factor $c = -1$.

But what happens outside the unit disk? Since all finite Blaschke products, and in particular the partial products of (5.54), satisfy the symmetry relation $f\big(1/\overline{z}\big) = 1/\overline{f(z)}$, we can use this identity to study infinite Blaschke products in the exterior $\mathbb{E} : \mathbb{C} \setminus \overline{\mathbb{D}}$ of the unit circle. With $Z := \{1/\overline{z}_k : \ k = m+1, m+2, \ldots\}$, the product (5.54) converges (absolutely and normally) in $\mathbb{E} \setminus Z$ to an analytic function f. Setting $f(z) := \infty$ for $z \in Z$, we obtain an analytic (meromorphic) function in all of $\mathbb{E}$, having poles at the points of Z.

Finally, if $z_0 \in \mathbb{T}$ there are two possibilities: either z_0 is an accumulation point of zeros z_k, or the distance of z_0 to zero set is positive. In the first case there is no chance to extend the Blaschke product analytically to z_0, because any such extension must vanish in a neighborhood of z_0 by the identity theorem. In the second case the Blaschke product converges (absolutely and normally) in a neighborhood of z_0. To verify this we may assume that $z_0 = 1$ and estimate

$$\left|1 - \frac{|z_k|}{z_k}\,\frac{z_k - z}{1 - \overline{z}_k\, z}\right| = (1 - |z_k|)\left|\frac{z_k + |z_k|\, z}{(1 - \overline{z}_k\, z)\, z_k}\right| \le (1 - |z_k|)\,\frac{1 + |z|}{|1 - \overline{z}_k\, z|}. \tag{5.55}$$

Since $1/z_0 = \overline{z_0} \in \mathbb{T}$ is not an accumulation point of the points $\overline{z}_k$, we have $|1 - \overline{z}_k\, z| \ge 1/C > 0$ for all z in some neighborhood of z_0 and all $k = 0, 1, 2, \ldots$, so that the denominator on the right-hand side of (5.55) is bounded from below by a positive constant. Now the Blaschke condition implies the desired result.

While the Blaschke product depicted on the left-hand side of Figure 5.11 converges nowhere on $\mathbb{T}$ (all points on $\mathbb{T}$ are accumulation points of the zeros z_k), the one on the right-hand side converges at all points of $\mathbb{T}$, except at the tenth roots of unity.

5.5 Cauchy Integrals

The integral which appears in the Cauchy integral formula does not only allow one to investigate a *given* analytic function, it also yields a *constructive principle* for analytic functions: when the function f in the integrand is replaced by an arbitrary continuous function φ on the trace of γ, it defines a function f,

$$f(z) := \frac{1}{2\pi \mathrm{i}} \int_\gamma \frac{\varphi(w)}{w - z}\, dw, \qquad z \in G := \mathbb{C} \setminus [\gamma]. \tag{5.56}$$

According to Definition 4.2.19, the integral on the right-hand side of (5.56) (and sometimes also the function f) is referred to as the *Cauchy integral with density* φ along γ. The function $C(z, w) := 1/(w - z)$ is called the *Cauchy kernel.*

General Assumptions. In contrast to the Cauchy integrals encountered in the preceding section, here we do not in general assume that the path γ is closed.

In order to make the integral (5.56) well defined we assume for simplicity that the path γ is piecewise smooth, and that the density φ is a continuous function on the trace of $[\gamma]$. Note that there is a well-established theory of Cauchy integrals for more general classes of densities and less regular paths.

Definition of the Density. In Definition 5.56 we follow the usual convention that the density φ is prescribed *on the trace* $[\gamma]$ of the path. However, it would be more natural to transplant it from the *parameter interval* $[\alpha, \beta]$ of γ, so that the composition $\varphi \circ \gamma$ in the definition of the path integral along γ is replaced by a function ψ on $[\alpha, \beta]$.

To explain the difference between the two concepts, we say that $z_0 \in [\gamma]$ is a *simple point* of γ, if there is only one $t \in [\alpha, \beta]$ with $z_0 = \gamma(t)$, otherwise z_0 is called a *multiple point* of γ.

Defining the density on the parameter interval is more flexible, because it allows that φ attains different values at the various "branches" of $[\gamma]$ at multiple points of γ. As mentioned above, we keep to the traditional concept, but all results are formulated such that they can easily be interpreted in the more general setting. The fundamental result on Cauchy integrals is Theorem 4.2.20, which tells us that the function f defined by (5.56) is *analytic* in $D := \mathbb{C} \setminus [\gamma]$ and *tends to zero at infinity.*

Before we study further properties of Cauchy integrals, let us consider two typical examples to get some feeling for how such functions look like.

Example 5.5.1. Let γ be the standard parametrization of the segment $[-1, 1]$ on the real axis. The Cauchy integral with constant density 1 along γ can be computed explicitly,

$$f(z) := \frac{1}{2\pi \mathrm{i}} \int_\gamma \frac{dw}{w-z} = \frac{1}{2\pi \mathrm{i}} \operatorname{Log} \frac{z-1}{z+1}.$$

Here, as usual, Log stands for the principal value of the logarithm. Indeed, if for example $\operatorname{Im} z > 0$, then $\operatorname{Im}(w-z) < 0$ for all $w \in [-1,1]$, so that $w \mapsto \operatorname{Log}(w-z)$ is an analytic branch of the logarithm and a primitive of $w \mapsto 1/(w-z)$, and the fundamental theorem of complex calculus (Theorem 4.2.11) yields

$$\int_\gamma \frac{dw}{w-z} = \operatorname{Log}(1-z) - \operatorname{Log}(-1-z) = \operatorname{Log} \frac{z-1}{z+1}. \tag{5.57}$$

The other cases can be treated similarly. Note that the last equality in (5.57) is *not automatic* but has to be verified (and is left as an exercise); in general we *cannot be sure* that $\operatorname{Log} a - \operatorname{Log} b = \operatorname{Log}(a/b)$.

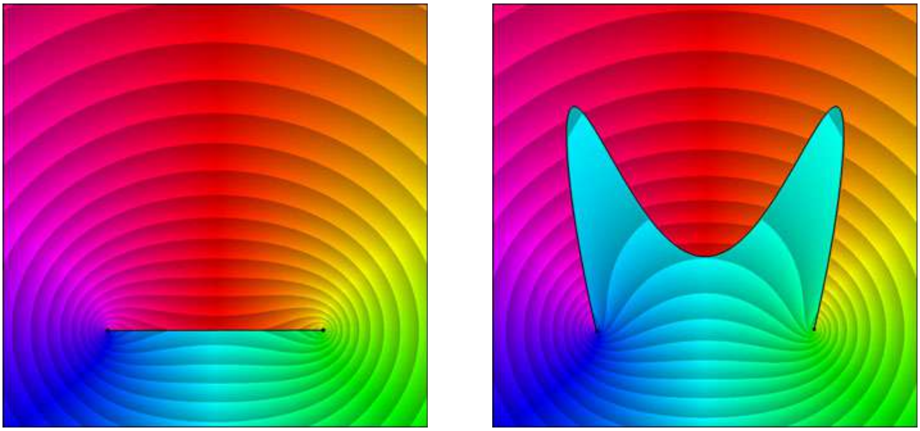


Figure 5.12: Values of two Cauchy integrals with constant density 1

The phase portrait of the function f in some square D is shown Figure 5.12 (left). The picture on the right visualizes what happens if the path $\gamma_0 = [-1, 1]$ of integration is deformed (homotopic) to another path γ_1 with the same endpoints. We see that the function has changed only in the (blue) region B bounded by $[\gamma_0]\cup[\gamma_1]$ which was swept through by the moving path. How can one explane this ?

First of all we remark that the density $\varphi = 1$ is not just defined on $[\gamma]$, but is an analytic function on the entire plane, so the integrand of the Cauchy integral (5.56) is analytic in $\mathbb{C}\setminus\{z\}$. Furthermore, for all points $z \notin B$, the paths γ_0 and γ_1 are homotopic in $\mathbb{C}\setminus\{z\}$, and Cauchy's integral theorem (Theorem 4.2.32) tells us that for *these points* the Cauchy integrals along γ_0 and γ_1 attain the same value.

There is still another interpretation of the two pictures in Figure 5.12. Once we know that the analytic functions f_0 and f_1, corresponding to integration along γ_0 and γ_1 respectively, coincide on some disk, say D_0, both can be considered as *analytic continuations* of the function element $(f_0, D_0) = (f_1, D_0)$ to the respective domains $D \setminus [\gamma_0]$ and $D \setminus [\gamma_1]$ (recall that D denotes the square shown in Figure 5.12). So what we see in the picture on the right is just *another analytic branch* of the logarithm $\log{(z-1)/(z+1)}$.

Piecewise Analytic Functions. The two windows of Figure 5.13 depict phase portraits of Cauchy integrals f with some more involved density functions. Since the densities have no special meaning, we do not describe them by explicit formulas. Functions like in Figure 5.13 (right) are often called *piecewise analytic*.

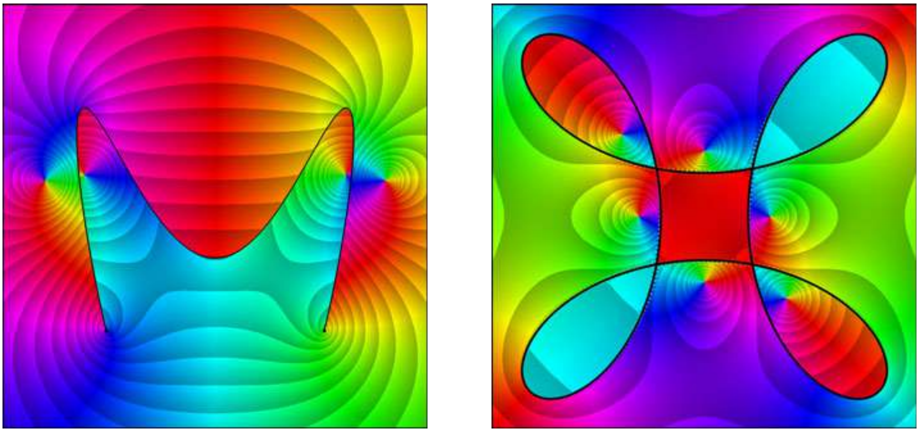


Figure 5.13: Two functions defined by Cauchy integrals with variable density

Behavior Near the Trace of the Path. In all examples we observe that the limits of the functions f on both sides of $[\gamma]$ are different, the functions f suffer a jump along the trace of γ. We shall show that this behavior is typical for Cauchy integrals.

Our first result serves to split the Cauchy integral into a regular part and a main part; the latter dominates the behavior near the path of integration. Some regularity assumptions on the path γ and the density φ are given in the next definition.

Definition 5.5.1. Let $\gamma : [\alpha, \beta] \to \mathbb{C}$ be continuously differentiable. A point z_0 on $[\gamma]$ is called *regular*, if $\gamma'(t_0) \neq 0$ for all $t_0 \in [\alpha, \beta]$ with $\gamma(t_0) = z_0$. If $\gamma'(t) \neq 0$ for all $t \in [\alpha, \beta]$, we say that γ is a *regular path*. A density $\varphi : [\gamma] \to \mathbb{C}$ is said to be *continuously differentiable along* γ, if $\varphi \circ \gamma$ is continuously differentiable on $[\alpha, \beta]$.

Lemma 5.5.2 (Fundamental Lemma)**.** *Let the density φ be continuously differentiable along a smooth path γ. If z_0 is a regular point on $[\gamma]$, then the function ψ, defined by*

$$\psi(z) := \int_\gamma \frac{\varphi(w) - \varphi(z_0)}{w - z}\, dw, \qquad z \in G := \mathbb{C} \setminus [\gamma], \tag{5.58}$$

has a finite limit at z_0.

Proof. 1. We assume that γ is defined on $[0, 1]$. In the first part of the proof we suppose that z_0 is a *simple point* of γ, i.e., there is exactly one $t_0 \in [0, 1]$ with $\gamma(t_0) = z_0$. To begin with, we consider the case where z_0 is the initial point of γ.

2. Since the functions γ and $\varphi \circ \gamma$ are continuously differentiable, they can be represented on $[0, 1]$ by

$$\gamma(t) = a_0 + t\,\big(a_1 + g(t)\big), \quad \varphi\big(\gamma(t)\big) = b_0 + t\,\big(b_1 + f(t)\big), \tag{5.59}$$

where $a_0 := \gamma(0) = z_0$, $a_1 := \gamma'(0) \neq 0$, $b_0 := \varphi(z_0)$, and $b_1 := (\varphi \circ \gamma)'(0)$. The functions f and g are continuous on $[0, 1]$, continuously differentiable on $(0, 1]$, and satisfy $f(0) = g(0) = 0$.

3. Inserting (5.59) in the definition of ψ, we obtain a natural extension of ψ to z_0,

$$\psi(z_0) := \int_\gamma \frac{\varphi(w) - \varphi(z_0)}{w - z_0}\, dw = \int_0^1 \frac{b_1 + f(t)}{a_1 + g(t)}\, dt. \tag{5.60}$$

Due to the assumption that z_0 is simple, we have $t(a_1 + g(t)) = \gamma(t) - z_0 \neq 0$ for all $t \in (0, 1]$. Since also $a_1 + g(0) \neq 0$, the integrand is a continuous function on $[0, 1]$, so that there exists an M with

$$\left| \frac{\varphi(w) - \varphi(z_0)}{w - z_0} \right| \leq M, \qquad w \in [\gamma] \setminus \{z_0\}. \tag{5.61}$$

4. In order to show continuity of ψ at z_0, we fix a positive ε and verify the classical ε-δ-definition. The proof is based on the representation

$$\psi(z) - \psi(z_0) = \int_\gamma \frac{z - z_0}{w - z}\, \frac{\varphi(w) - \varphi(z_0)}{w - z_0}\, dw, \qquad w \in G, \tag{5.62}$$

and will be achieved in several steps.

5. After an appropriate rotation of the complex plane we may assume that $\gamma'(0)$ is positive. Let K and L denote the two sectors with vertex z_0 defined by

$$K := \{z \in \mathbb{C} \setminus \{z_0\} : |\mathrm{Arg}\,(z_0 - z)| < 2\pi/3\}, \tag{5.63}$$
$$L := \{z \in \mathbb{C} \setminus \{z_0\} : |\mathrm{Arg}\,(z - z_0)| < \pi/6\}. \tag{5.64}$$

A short excursion to elementary geometry shows that for all $z \in K$ and $w \in L$

$$\left| \frac{z - z_0}{w - z} \right| \le 2, \qquad z \in K,\ w \in L. \tag{5.65}$$

6. The assumptions on γ (in particular $\gamma'(0) > 0$) and the construction of L ensure the existence of a positive number τ such that $\gamma(t) \in L$ for all $t \in (0, \tau]$. We fix one such τ, which additionally satisfies $\tau < \varepsilon/(4M \max|\gamma'|)$, and denote the restrictions of γ to the intervals $[0, \tau]$ and $[\tau, 1]$ by γ_0 and γ_1, respectively. Then we get, using (5.61) and (5.65),

$$\left| \int_{\gamma_0} \frac{z - z_0}{w - z} \, \frac{\varphi(w) - \varphi(z_0)}{w - z_0} \, dw \right| \le 2\, M\, \tau < \frac{\varepsilon}{2}, \qquad z \in K. \tag{5.66}$$

7. Since z_0 has a positive distance from $[\gamma_1]$, we find positive constants ϱ and C such that

$$|w - z_0| \ge 1/C, \quad |w - z| \ge 1/C, \qquad w \in [\gamma_1],\ z \in K,\ |z - z_0| < \varrho. \tag{5.67}$$

Choosing δ so small such that $\delta < \varepsilon/(4\, C^2 \max|\varphi|)$ and $0 < \delta < \varrho$ we get

$$\left| \int_{\gamma_1} \frac{z - z_0}{w - z} \, \frac{\varphi(w) - \varphi(z_0)}{w - z_0} \, dw \right| < \frac{\varepsilon}{2}, \qquad |z - z_0| < \delta. \tag{5.68}$$

8. Since $\gamma = \gamma_0 \oplus \gamma_1$, we infer from (5.62), (5.66) and (5.68) that

$$|\psi(z) - \psi(z_0)| \le \varepsilon, \qquad z \in K,\ |z - z_0| < \delta. \tag{5.69}$$

Before we continue with obtaining a corresponding estimate *without the restriction* $z \in K$, we briefly discuss the case where z_0 is not an endpoint of γ. Considering $z_0 = \gamma(t_0)$ as a common endpoint of the restrictions of γ to $[0, t_0]$ and $[t_0, 1]$, respectively, we see that the estimate (5.69) remains in force for such points z_0 if the sector K defined in (5.63) is replaced by the double sector

$$K := \{z \in \mathbb{C} \setminus \{0\} : \pi/3 < |\mathrm{Arg}\,(z_0 - z)| < 2\pi/3\}. \tag{5.70}$$

9. In order to get rid of the condition $z \in K$, we vary the point z_0 along the trace of γ. Since then the function ψ depends on two variables z and z_0, we denote it by $\Psi(z, z_0)$, i.e.,

$$\Psi(z, z_0) := \int_\gamma \frac{\varphi(w) - \varphi(z_0)}{w - z} \, dw, \qquad z_0 \in [\gamma],\ z \in G \cup \{z_0\} \tag{5.71}$$

with the modified definition (5.62) if $z = z_0$. In order to show that the function $z_0 \mapsto \Psi(z_0, z_0)$ is continuous, we remark that all parameters in Steps 2 to 7 can be chosen *independently* of z_0, so that we get uniform estimates

$$|\Psi(z, z_0) - \Psi(z_0, z_0)| < \varepsilon, \qquad z_0 \in \gamma,\ z \in K(z_0),\ |z - z_0| < \delta, \tag{5.72}$$

where $K(z_0)$ denotes the (double) sector associated with z_0, rotated according to the argument of $\gamma'(t_0)$ (see Figure 5.14).

If δ is sufficiently small, then for all points z_0 and z_1 on $[\gamma]$ with $|z_0 - z_1| < \delta/2$ there exists a point z which belongs to $K(z_0)$ and $K(z_1)$ and satisfies $|z - z_0| < \delta$, $|z - z_1| < \delta$. For any such point z,

$$\big|\Psi(z, z_0) - \Psi(z_0, z_0)\big| < \varepsilon, \tag{5.73}$$
$$\big|\Psi(z, z_1) - \Psi(z_1, z_1)\big| < \varepsilon. \tag{5.74}$$

Moreover, the points z can be chosen such that their distance to $[\gamma]$ is bounded by a positive constant, uniformly for all z_0 and z_1.

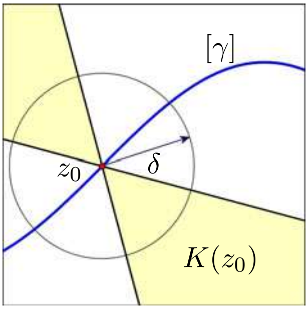


Figure 5.14: The sector $K(z_0)$

Decreasing δ, if necessary, the standard estimate for the integral (5.71) shows that also

$$\big|\Psi(z, z_0) - \Psi(z, z_1)\big| < \varepsilon, \qquad |z_0 - z_1| < \delta/2. \tag{5.75}$$

Using the triangle inequality, we conclude from (5.73), (5.74) and (5.75) that

$$\big|\Psi(z_1, z_1) - \Psi(z_0, z_0)\big| < 3\,\varepsilon, \qquad z_0, z_1 \in [\gamma],\ |z_0 - z_1| < \delta/2, \tag{5.76}$$

which proves the continuity of $z_0 \mapsto \Psi(z_0, z_0)$ on γ.

10. In order to complete the proof for simple points z_0, we consider any point $z \in G$ with $|z - z_0| < \delta/2$. Then there exists a point z_1 on $[\gamma]$ such that $z \in K(z_1)$, $|z_1 - z_0| < \delta/2$, and $|z - z_1| < \delta$, and from (5.73), (5.74) and (5.76) we get that

$$\big|\Psi(z, z_1) - \Psi(z_1, z_1)\big| < \varepsilon, \qquad \big|\Psi(z_1, z_1) - \Psi(z_0, z_0)\big| < 3\,\varepsilon$$

and finally

$$|\psi(z) - \psi(z_0)| = \big|\Psi(z, z_0) - \Psi(z_0, z_0)\big| < 4\,\varepsilon, \qquad z \in G,\ |z - z_0| < \delta/2.$$

11. Now let z_0 be not necessarily a simple point. If the number of points $t \in [0, 1]$ with $\gamma(t) = z_0$ were infinite, there would exist a sequence (t_k) of such points converging to $t_0 \in [0, 1]$. By continuity of γ we have $\gamma(t_0) = z_0$, and since z_0 is regular, $\gamma'(t_0) \neq 0$. Because γ' is continuous, there exists a constant $c \in \mathbb{C}$ such that $\operatorname{Re}\big(c\gamma'(t)\big) > 0$ for all $t \in [0, 1]$ sufficiently close to t_0. But this contradicts

$$\int_{t_0}^{t_k} c\,\gamma'(t)\,dt = c\,\big(\gamma(t_k) - \gamma(t_0)\big) = c\,\big(z_0 - z_0\big) = 0$$

for all k. Consequently there are only finitely many points $t_1, \dots, t_n \in [0,1]$ with $\gamma(t_k) = z_0$. The assertion then follows on applying the considerations of the preceding steps repeatedly, with t_0 replaced by t_k for $k = 1, \dots, n$. $\square$

Behavior Near Endpoints of the Path. Our next concern is to investigate the behavior of a Cauchy integral f near an endpoint of the integration path. Roughly speaking, we shall show that f has a *logarithmic singularity* at these points.

We begin by remarking that $G := \mathbb{C} \setminus [\gamma]$ is an open set which consists of (at most countably many) connected components. On each of these components of G, the Cauchy integral with density one defines a special analytic branch of the logarithm $\log\big((b-z)/(a-z)\big)$.

Definition 5.5.3. The function L_γ defined by

$$L_\gamma(z) = \log_\gamma \frac{b-z}{a-z} := \int_\gamma \frac{dw}{w-z}, \qquad z \in \mathbb{C} \setminus [\gamma] \tag{5.77}$$

is referred to as the *canonical logarithm* of the path γ.

Which of the two equivalent expressions in (5.77) we prefer will depend on the context. Denoting by $w \mapsto \arg_\gamma(w-z)$ a continuous branch of the argument of $w - z$ along γ, we get the representation

$$L_\gamma(z) = \operatorname{Log}|b-z| - \operatorname{Log}|a-z| + \mathrm{i}\,\arg_\gamma(b-z) - \mathrm{i}\,\arg_\gamma(a-z). \tag{5.78}$$

Note that L_γ is also well defined on the unbounded component of G and tends to zero as $z \to \infty$, so that it is natural to define $L_\gamma(\infty) := 0$. If γ is closed, then $L_\gamma(z)$ is constant on the connected components of G and equal to $2\pi\mathrm{i}\,\mathrm{wind}(\gamma, z)$.

Figures 5.15 and 5.16 show phase portraits of the canonical logarithm for several paths γ.

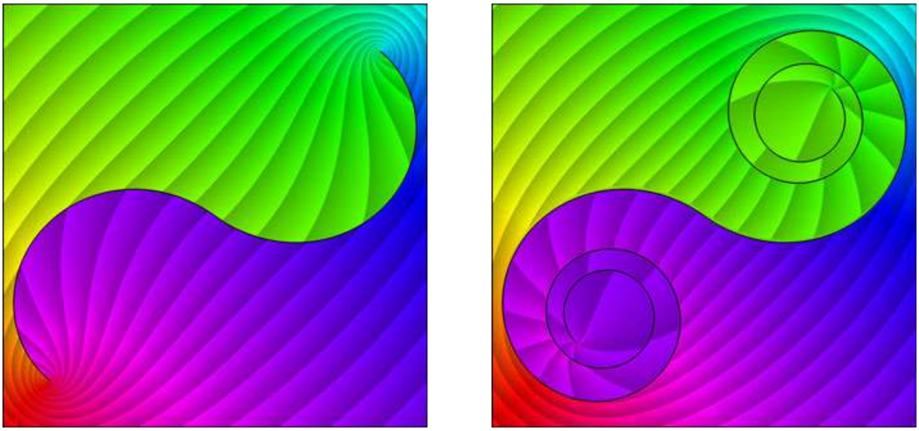


Figure 5.15: Phase portraits of L_γ for an evolving double-spiral path

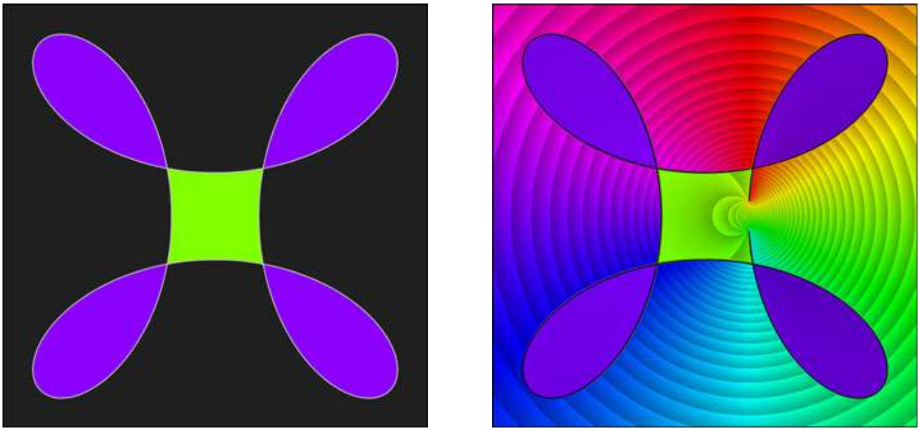


Figure 5.16: Phase portraits of L_γ for a closed and an almost closed path

Exploring the Canonical Logarithm. Let us devote some time to interpret and explore these pictures. In the *unbounded component* of G the canonical logarithm is just the analytic branch of $\log\big((z-b)/(z-a)\big)$ which vanishes at infinity. If γ is closed, this branch is identically zero, as in Figure 5.16 (left). Otherwise its phase portrait can be obtained from the portrait of

$$L^*(z) := \operatorname{Log}\big((z-1)/(z+1)\big)$$

by a similarity transformation (translation, rotation, dilation) which sends -1 to a and 1 to b. If the segment $[a,b]$ is short, this has the effect of “zooming out”, as if we were looking at the portrait of L^* from afar. The contour lines are then almost circular, as in the picture on the right in Figure 5.16.

Interestingly, the phase portrait in the unbounded component makes it easy to decide which endpoint is the initial point of γ: the ray from the terminal point b to the initial point a must end up in the red region (positive real values) of the phase portrait. For example, the spiral path of Figure 5.15 starts in the lower left corner and terminates in the upper right corner.

Departing from a point in the unbounded component of G, we can explore the rest. The *real part* of L_γ can easily be determined everywhere, it is just the logarithm of the distance ratio $|z-b|/|z-a|$. We only have to observe carefully the argument of $(z-b)/(z-a)$, which determines the *imaginary part* of L_γ. So we have to monitor how the *angle* between the rays from $z-b$ and $z-a$ is changing. If, for example, we travel into the right arm of the spiral shown in Figure 5.15, $\arg(z-b)$ increases monotonously, while $\arg(z-a)$ oscillates. So the imaginary part of $L_\gamma(z)$ is getting bigger and bigger. At the same time the real part of L_γ has a tendency to decrease, since we move closer to the point b, while the distance to a remains almost unchanged. Since the real part decays very slowly (logarithmically), the value of L_γ stays in a small sector near the upper part of the imaginary axis, the phase portrait of this spiral arm shows shades of green. Figure 5.17 shows the (circular) contour lines of the real part (left) and the imaginary part (right) of L_γ for the same spiral path as on the right of Figure 5.15. But why are these lines circular arcs?

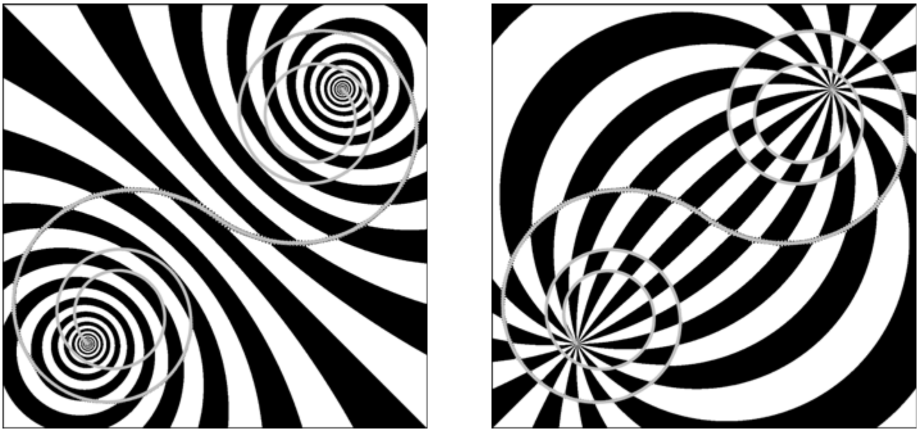


Figure 5.17: Contour lines of real and imaginary part of L_γ for a spiral path

When we *cross the trace* of the path, the value of L_γ jumps by an integer multiple of $2\pi\mathrm{i}$. At a *simple* point of γ this jump is either $2\pi\mathrm{i}$ or $-2\pi\mathrm{i}$, where the positive sign has to be taken when the path in front runs from left to right. This happens, for instance, if γ is closed and the winding number of the component which we are leaving is smaller than the winding number of the component we are entering.

Let us apply this to the picture on the right in Figure 5.15. If we enter one of the (purple) loops coming from the exterior domain, the path in front runs from right to left, and consequently the value of L_γ jumps by $-2\pi\mathrm{i}$.

But in fact we can say much more about the value of L_γ. Since we entered the loop coming from the exterior domain at a point which has almost the same distance to a and b, the value of L_γ was rather close to zero before, so that it must be close to $-2\pi\mathrm{i}$ after the border crossing. This explains the almost uniform blue color inside the loops.

A Representation of the Cauchy Integral. After these experiments it is time to state the main result of our investigations so far.

Theorem 5.5.4. *Let f be defined by the Cauchy integral (5.56) with continuously differentiable density φ along a smooth path γ from a to b. If z_0 is a regular point of γ, then the function*

$$f_0(z) := f(z) - \frac{\varphi(z_0)}{2\pi\mathrm{i}} \log_\gamma \frac{b-z}{a-z} \tag{5.79}$$

is continuous on $(\mathbb{C} \setminus [\gamma]) \cup \{z_0\}$.

Proof. The assertion is just a reformulation of Lemma 5.5.2, using the definition (5.77) of the special branch of the logarithm and the representation

$$2\pi\mathrm{i}\, f(z) = \int_\gamma \frac{\varphi(w)}{w-z}\,dw = \int_\gamma \frac{\varphi(w)-\varphi(z_0)}{w-z}\,dw + \varphi(z_0)\int_\gamma \frac{dw}{w-z}. \tag{5.80}$$

□

The crucial point about Theorem 5.5.4 is the continuity of f_0 at the point z_0 on γ. If z_0 is a regular *endpoint* of γ, then the modulus of $\log_\gamma\big((b-z)/(a-z)\big)$ tends to infinity as $z \to z_0$, and we obtain the following corollary.

Corollary 5.5.5. *If z_0 is a regular endpoint of γ and $\varphi(z_0) \neq 0$, then the asymptotic behavior of f at z_0 is given by*

$$f(z) \sim \frac{\varphi(z_0)}{2\pi\mathrm{i}} \log_\gamma \frac{b-z}{a-z}. \tag{5.81}$$

The Jump Relation. The most important consequence of Theorem 5.5.4 is a celebrated *jump relation* for Cauchy integrals, which generalizes our previous observations for integrals with constant density.

In order to keep formulations simple, we assume that γ is smooth, and z_0 is a simple regular point on $[\gamma]$. Furthermore we suppose that z_0 is not an endpoint of γ. Under these assumptions there exists a disk D centered at z_0 such that $D \setminus [\gamma]$ consists of two connected components D^- and D^+. The vector $\gamma'(t_0)$ is tangent to $[\gamma]$ at z_0 and we choose the $\pm$–notation so that $\mathrm{i}\,\gamma'(t_0)$ points into the interior of D^+ (walking along $[\gamma]$ in the direction of increasing t, the region D^+ lies to the left and D^- lies to the right). Finally, we define the *one-sided limits* of f at z_0 by

$$f^-(z_0) := \lim_{z\to z_0, z\in D^-} f(z), \qquad f^+(z_0) := \lim_{z\to z_0, z\in D^+} f(z). \tag{5.82}$$

Theorem 5.5.6 (Plemelj-Sokhotsky Jump Relation)**.** *Let f be defined by the Cauchy integral (5.56) with continuously differentiable density along a smooth path γ. Then, at every simple regular point z_0 of $[\gamma]$ which is not an endpoint, the one sided limits (5.82) exist and satisfy*

$$f^+(z_0) - f^-(z_0) = \varphi(z_0).$$

Proof. By virtue of Theorem 5.5.4, the difference $f^+(z_0) - f^-(z_0)$ is solely determined by the behavior of $\log_\gamma\big((b-z)(a-z)\big)$ in a neighborhood of z_0. It follows from (5.78) that this function jumps by $2\pi\mathrm{i}$ when z moves across $[\gamma]$ from the minus side to the plus side. This can be seen most easily by observing the winding numbers of the path which results from γ by complementing it to a closed path with a line from b to a. □

We point out that the statement of Theorem 5.5.6 is purely local. It can be generalized to Cauchy integrals along piecewise smooth paths $\gamma = \gamma_1 \oplus \ldots \oplus \gamma_n$, with densities that are continuous along γ and smooth on every component γ_k. To obtain the result at a common endpoint z_0 of two components γ_{k-1} and γ_k, Theorem 5.5.4 can be applied separately to γ_{k-1} and γ_k.

We also remark that the simplicity of the jump relation is the reason why the Cauchy integral is usually introduced with the factor $1/(2\pi\mathrm{i})$ in front.

The Cauchy Singular Integral. Our final concern is an option for computing the one sided limits $f^+(z_0)$ and $f^-(z_0)$ *directly* from φ on $[\gamma]$, without making the detour to the domain $\mathbb{C} \setminus [\gamma]$ and taking limits afterwards. Inserting z_0 directly in the definition (5.56) of the Cauchy integral is problematic, at least if the integral is understood as a Riemann integral, because then the integrand becomes infinite at $w = z_0$. Even the interpretation as an improper Riemann integral or a Lebesgue integral does not help much, since typically the integrand has a pole of first order at z_0.

Nevertheless, there is an appropriate modification of the integral concept which resolves the problem. The trick is to cut off two (almost) *symmetric pieces* of the path in a neighborhood of z_0 in order to eliminate the singularity, and then to consider what happens when the lengths of these two pieces tend to zero at the same rate. It turns out that, under certain conditions, the big values of the integrand cancel each other, so that a finite limit exists. The technical details of this procedure are described next.

Let $z_0 = \gamma(t_0)$ be a simple regular point on the smooth path $\gamma : [\alpha, \beta] \to \mathbb{C}$. Then, for any $\varepsilon > 0$, the restrictions of γ to $[\alpha, t_0 - \varepsilon]$ and $[t_0 + \varepsilon, \beta]$ are denoted γ_ε^- and γ_ε^+, respectively, and we set $\gamma_\varepsilon := \gamma_\varepsilon^- + \gamma_\varepsilon^+$. Note that $\gamma_\varepsilon^\pm$ may be empty.

Definition 5.5.7. Let γ be a smooth simple path. Then the *Cauchy singular integral* along γ with density φ is defined by

$$S\varphi(z_0) = \frac{1}{\pi\mathrm{i}} \,\text{v.p.} \int_\gamma \frac{\varphi(w)}{w - z_0}\, dw := \lim_{\varepsilon \to +0} \frac{1}{\pi\mathrm{i}} \int_{\gamma_\varepsilon} \frac{\varphi(w)}{w - z_0}\, dw, \tag{5.83}$$

for all points $z_0 \in [\gamma]$ where this limit exists.

Integrals which are defined by the limit construction in (5.83) are said to be *principal value integrals*, the abbreviation *v.p.* stands for “valeur principale”. By convention, the factor in front is $1/(\pi\mathrm{i})$ and not $1/(2\pi\mathrm{i})$.

Singular integrals form a large and fascinating subject with an elaborate theory and still ongoing research. In Definition 5.5.7 we did not worry about the conditions for the existence of the limit (5.83). In this book we consider a simple situation, where a sufficient criterion can be derived without too much machinery.

Lemma 5.5.8. *If φ is continuously differentiable along a two-times-continuously-differentiable path γ from a to b, then the limit $S\varphi(z)$ in (5.83) exists for all regular simple points $z \in [\gamma] \setminus \{a, b\}$.*

Remark 5.5.9. If the definition of the principle value integral is modified such that symmetric neighborhoods of t_k are excluded from the parameter interval for *all* points t_k with $\gamma(t_k) = z_0$, the assumption of Lemma 5.5.8 that z is simple can be omitted.

Proof. Let $\gamma : [\alpha, \beta] \to \mathbb{C}$, and let $z_0 := \gamma(t_0)$ with $t_0 \in (\alpha, \beta)$. By virtue of the assumptions on γ and φ, we can represent γ, γ', and $\varphi \circ \gamma$ as

$$\begin{aligned} \gamma(t_0 + t) \quad &= \gamma_0 + t\,\big(\gamma_1 + g(t)\big), && (5.84)\\ \gamma'(t_0 + t) \quad &= \gamma_1 + g(t) + t\,g'(t), && (5.85)\\ \varphi\big(\gamma(t_0 + t)\big) &= \varphi_0 + t\,\big(\varphi_1 + h(t)\big) && (5.86) \end{aligned}$$

with $\gamma_0 := \gamma(t_0)$, $\gamma_1 := \gamma'(t_0) \neq 0$, $\varphi_0 := \varphi(z_0)$, $\varphi_1 := (\varphi \circ \gamma)'(t_0)$. The definitions $h(0) := 0$, $g(0) = 0$, and $g'(0) := \gamma''(t_0)/2$, extend h to a continuous and g to a continuously differentiable function on $[\alpha - t_0, \beta - t_0]$. Omitting the factor $1/(\pi \mathrm{i})$ in the definition of the singular integral, we set

$$S_\varepsilon \varphi(z_0) := \int_{\gamma_\varepsilon(z_0)} \frac{\varphi(w)}{w - z_0}\, dw.$$

For $0 < \varepsilon_1 < \varepsilon_2 < \varepsilon$, we consider the difference

$$S_{\varepsilon_1}(z_0) - S_{\varepsilon_2}(z_0) = \int_{t_0-\varepsilon_2}^{t_0-\varepsilon_1} \frac{\varphi(\gamma(t))}{\gamma(t) - \gamma(t_0)}\, \gamma'(t)\, dt + \int_{t_0+\varepsilon_1}^{t_0+\varepsilon_2} \frac{\varphi(\gamma(t))}{\gamma(t) - \gamma(t_0)}\, \gamma'(t)\, dt.$$

Substituting the integration variable t by $t_0 - t$ in the first, and by $t_0 + t$ in the second integral, and inserting (5.84), (5.85), (5.86), we end up with a representation of $S_{\varepsilon_1}(z_0) - S_{\varepsilon_2}(z_0)$ as

$$\int_{\varepsilon_1}^{\varepsilon_2} \frac{\varphi_0\,\big(\gamma_1 + g(t)\big)\,\big(\gamma_1 + g(-t)\big) - \varphi_0\,\big(\gamma_1 + g(-t)\big)\,\big(\gamma_1 + g(t)\big) + t\,r(t)}{t\,\big(\gamma_1 + g(t)\big)\big(\gamma_1 + g(-t)\big)}\, dt,$$

where r is a continuous function on $[\alpha - t_0, \beta - t_0]$. Deleting all terms which compensate each other, just $t\,r(t)$ remains in the numerator. After cancelling t we see that the denominator does not vanish if ε is sufficiently small, since $\gamma_1 \neq 0$ and $g(0) = 0$. So we arrive at an integral of a *continuous* function on the interval $[\varepsilon_1, \varepsilon_2]$. By the standard integral estimate, $S_{\varepsilon_1}(z_0) - S_{\varepsilon_2}(z_0)$ becomes arbitrarily small when ε tends to zero. Consequently, by Cauchy's criterion, the limit of $S_\varepsilon(z_0)$ for $\varepsilon \to +0$ exists. □

Theorem 5.5.10. *Let f be defined by the Cauchy integral with a continuously differentiable density φ along a two-times-continuously-differentiable path γ from a to b. Then, for all regular simple points $z \in [\gamma] \setminus \{a, b\}$,*

$$f^+(z) = \frac{1}{2}\,\big(S\varphi(z) + \varphi(z)\big), \qquad f^-(z) = \frac{1}{2}\,\big(S\varphi(z) - \varphi(z)\big).$$

Proof. 1. We start with the constant density function $\varphi = 1$ and study the limit values of

$$(S1)(z) := \frac{1}{2\pi \mathrm{i}} \int_\gamma \frac{dw}{w-z}, \qquad z \in \mathbb{C} \setminus [\gamma].$$

To avoid wasting time with complicated computation, let us look again at Figure 5.12 from which we can learn a useful trick. Assume that we are interested in a limit value of the function f at a point z_0 on the (black) interval from below. Then all that we need to do is to '*lift the curtain*' as shown in the picture on the right. The function is not changed by the deformation of the integration path in the region below the interval, but now z_0 lies in the interior of the domain! By lifting the curtain we get immediate access to the limit values which were hidden so far.

In order to exploit this idea for actually *computing* the limit values of $S1$, say on the "plus side", we replace a short (almost symmetric) piece of the path γ in a neighborhood of z_0 by a small semicircle in the "minus side" of the path. Then the limit of the Cauchy integral at z_0 on the "plus side" is equal to the value at z_0 of the Cauchy integral along the modified path. If we contract the semicircle to the point z_0, the integral along that semicircle converges to $\pi\mathrm{i}$, while the integral along the unchanged part of γ converges (per definition) to (one half of) the *singular* Cauchy integral. Now let us check more formally how this procedure works.

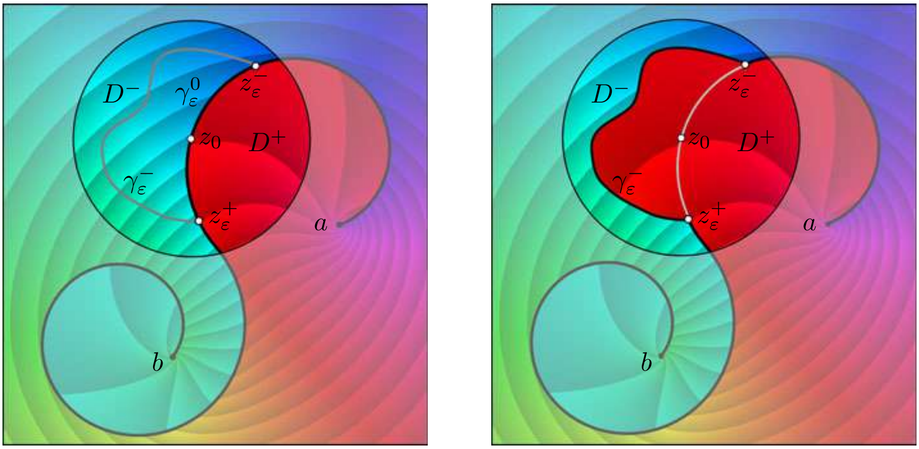


Figure 5.18: Deformation of a path for computing one-sided limit values

Let $z_0 = \gamma(t_0)$ be a regular simple point on $[\gamma] \setminus \{a, b\}$. We choose a (small) disk D centered at z_0 such that $D \setminus [\gamma]$ consists of two connected components D^- and D^+ (see (5.82)). Then, for all sufficiently small positive ε the points $z_\varepsilon^- := \gamma(t_0 - \varepsilon)$ and $z_\varepsilon^+ := \gamma(t_0 + \varepsilon)$ lie in D. We connect these points by a path γ_ε^- from z_ε^- to z_ε^+ in D^- (see Figure 5.18 left) and replace the portion γ_ε^0 of γ between z_ε^- and z_ε^+ by γ_ε^-. Abusing notation slightly, we write the modified path γ_ε in the intuitive form $\gamma_\varepsilon := \gamma - \gamma_\varepsilon^0 + \gamma_\varepsilon^-$. By Cauchy's integral theorem, the function $(S1)_\varepsilon$ defined by

$$(S1)_\varepsilon(z) := \frac{1}{2\pi\mathrm{i}} \int_{\gamma_\varepsilon} \frac{dw}{w-z}, \qquad z \in \mathbb{C} \setminus [\gamma_\varepsilon],$$

coincides with $(S1)$ in D_+. Consequently, the limit $(S1)^+(z_0)$ exists and is equal to the value $(S1)_\varepsilon(z_0)$. In order to calculate it, we first write it in a convenient form

$$(S1)_\varepsilon(z_0) = \frac{1}{2\pi\mathrm{i}} \int_{\gamma_\varepsilon} \frac{dw}{w-z_0} = \frac{1}{2\pi\mathrm{i}} \int_{\gamma-\gamma_\varepsilon^0} \frac{dw}{w-z_0} + \frac{1}{2\pi\mathrm{i}} \int_{\gamma_\varepsilon^-} \frac{dw}{w-z_0}.$$

The first integral on the right-hand side appears in the definition of the principal value integral, and it converges to $(1/2)(S1)(z_0)$ as $\varepsilon \to 0$. The value of the last integral can be calculated explicitly,

$$\int_{\gamma_\varepsilon^-} \frac{dw}{w-z_0} = \log\big(\gamma(t_0+\varepsilon) - z_0\big) - \log\big(\gamma(t_0-\varepsilon) - z_0\big), \tag{5.87}$$

where log is a continuous branch of the logarithm along γ_ε^-. To find the limit for $\varepsilon \to +0$ we observe that

$$\frac{|\gamma(t_0+\varepsilon) - z_0|}{|\gamma(t_0-\varepsilon) - z_0|} \to 1, \quad \arg(\gamma(t_0+\varepsilon) - z_0) - \arg(\gamma(t_0-\varepsilon) - z_0) \to \pi,$$

and thus the limit of (5.87) is equal to $\pi\mathrm{i}$. Summarizing we have

$$(S1)^+(z_0) = \lim_{\varepsilon\to+0} (S1)_\varepsilon(z_0) = \frac{1}{2}\,(S1)(z_0) + \frac{1}{2}. \tag{5.88}$$

Replacing the arc γ_ε^- by an arc γ_ε^+ contained in D^+, we obtain in a similar manner

$$(S1)^-(z_0) = \lim_{\varepsilon\to+0} (S1)_\varepsilon(z_0) = \frac{1}{2}\,(S1)(z_0) - \frac{1}{2}. \tag{5.89}$$

2. In the general case we split f into the function ψ defined by (5.58), which is continuous at z_0 by Lemma 5.5.2, and a Cauchy integral with constant density,

$$f(z) = \frac{1}{2\pi\mathrm{i}} \int_\gamma \frac{\varphi(w) - \varphi(z_0)}{w-z}\, dw + \frac{\varphi(z_0)}{2\pi\mathrm{i}} \int_\gamma \frac{dw}{w-z}. \tag{5.90}$$

The integrand $w \mapsto \big(\varphi(w) - \varphi(z_0)\big)/\big(w-z\big)$ has a finite limit at z_0, so that $\psi(z_0)$ can be represented as

$$\psi(z_0) = \frac{1}{2\pi\mathrm{i}} \lim_{\varepsilon\to0} \int_{\gamma_\varepsilon} \frac{\varphi(w) - \varphi(z_0)}{w-z_0}\, dw = \frac{1}{2} S\varphi(z_0) - \frac{1}{2}\,\varphi(z_0)\,(S1)(z_0). \tag{5.91}$$

Combining this with (5.90) and the results (5.88), (5.89) of the first step, we finally get the desired result

$$2\, f^\pm(z_0) = S\varphi(z_0) - \varphi(z_0)\,(S1)(z_0) + \varphi(z_0)\,(S1)(z_0) \pm \varphi(z_0) = S\varphi(z_0) \pm \varphi(z_0).$$

□

The method of "lifting the curtain" works as long as the density φ is the restriction of an analytic function in a neighborhood of $[\gamma]$. By the way, this is one motivation for defining the density on the trace of γ and not on its parameter interval. If this is not so, we cannot deform the path without affecting the values of the Cauchy integral in the entire plane.

Cauchy integrals play an important role in the investigation of *boundary value problems*, in particular in the theory of linear and non-linear *Hilbert* and *Riemann-Hilbert problems*. We shall come back to this issue in Volume 2.

5.6 Integrals with Parameters

In this section we shall introduce and study a constructive method for building analytic functions which can be considered as a generalization of infinite series. While a series is a superposition of a countable number of summands, we are now concerned with superpositions of (analytic) functions which depend continuously on a parameter, which leads to the concept of *integrals with parameters* or *parameter integrals* in short.[2]

One prototype of a parameter integral we have already met is the Cauchy integral

$$f(z) = \frac{1}{2\pi \mathrm{i}} \int_\gamma \frac{\varphi(w)}{w - z}\, dw.$$

In the above interpretation, the function f is a *weighted superposition* of analytic functions $z \mapsto 1/(w - z)$. The family of these functions is parameterized by the point w on the trace of γ, or, more naturally, by the parameter t from the parameter interval of γ. The "weight" of a function $z \mapsto 1/(w-z)$ in the superposition is the value $\varphi(w)$ of the density.

With this interpretation in mind, it becomes intuitively clear, why a Cauchy integral defines an analytic function, no matter whether the density is analytic or not. Moreover, the distinguished role of the Cauchy kernel seems not to be relevant at all – so let us look what happens if we replace it by any (continuous) family of analytic functions.

After these considerations the following result is probably not a big surprise.

---

[2] To convince you of the utility of this technique we quote from *Surely you're joking, Mr. Feynman* p.86–87: 'This book [which is *Advanced Calculus* by Frederick Woods] also showed how to differentiate parameters under the integral sign - it's a certain operation. It turns out that's not taught very much in the universities; they don't emphasize it. But I caught on how to use this method and I used that one damn tool again and again. So because I was self-taught using that book, I had peculiar methods of doing integrals.

The result was, when guys at MIT or Princeton had trouble doing a certain integral, it was because they couldn't do it with the standard methods they had learned in school. If it was contour integration, they would have found it; if it was a simple series expansion, they would have found it. Then I come along and try differentiating under the integral sign, and often it worked. So I got a great reputation for doing integrals, only because my box of tools was different from everybody else's . . . '.

Theorem 5.6.1. *Let* $D \subset \mathbb{C}$ *be an open set, let* $I := [\alpha, \beta] \subset \mathbb{R}$ *be a (parameter) interval, and let* $g : I \times D \to \mathbb{C}$ *be a continuous function. If* g *is analytic with respect to* z*, i.e., the mappings* $z \mapsto g(t, z)$ *are analytic for every* $t \in I$*, then the function* $f : D \to \mathbb{C}$ *defined by*

$$f(z) := \int_\alpha^\beta g(t, z)\, dt \tag{5.92}$$

is analytic in D*. Furthermore, the derivative* $\partial g/\partial z$ *is continuous on* $I \times D$ *and*

$$f'(z) = \int_\alpha^\beta \frac{\partial g}{\partial z}(t, z)\, dt. \tag{5.93}$$

Proof. 1. First of all note that f is continuous in D. For verification, fix $z_0 \in D$, choose a closed disk $K_0 \subset D$ centered at z_0, and estimate $|f(z) - f(z_0)|$ by the standard integral estimate, using the uniform continuity of g on $I \times K_0$.

2. In order to prove that f is analytic, we pick $z_0 \in D$ and denote by D_0 the interior of the disk K_0 defined in the first step. Denoting by γ the standard parametrization of the boundary of D_0, Cauchy's integral formula tells us that

$$g(t, z) = \frac{1}{2\pi\mathrm{i}} \int_\gamma \frac{g(t, w)}{w - z}\, dw, \qquad t \in I,\ z \in D_0. \tag{5.94}$$

Integrating (5.94) along the parameter interval I, and interchanging the order of integration using Fubini's theorem, we obtain

$$f(z) = \frac{1}{2\pi\mathrm{i}} \int_I \int_\gamma \frac{g(t, w)}{w - z}\, dw\, dt = \frac{1}{2\pi\mathrm{i}} \int_\gamma \frac{f(w)}{w - z}\, dw, \quad z \in D_0. \tag{5.95}$$

Formula (5.95) represents f as a Cauchy integral with continuous density, so that f is analytic in D_0 by Theorem 4.2.20. Since z_0 was chosen arbitrarily, f is analytic in D.

3. Now applying Cauchy's integral formula for the derivative of f, we get

$$f'(z) = \frac{1}{2\pi\mathrm{i}} \int_\gamma \frac{f(w)}{(w - z)^2}\, dw = \frac{1}{2\pi\mathrm{i}} \int_\gamma \int_I \frac{g(t, w)}{(w - z)^2}\, dt\, dw, \quad z \in D_0.$$

Since the integrand is continuous on $I \times [\gamma]$, we can interchange the order of integration again, which results in

$$f'(z) = \frac{1}{2\pi\mathrm{i}} \int_I \int_\gamma \frac{g(t, w)}{(w - z)^2}\, dw\, dt = \int_I \frac{\partial g}{\partial z}(t, z)\, dt, \quad z \in D_0.$$

The last equality follows from Cauchy's integral formula applied to the mappings $z \mapsto g(t, z)$,

$$\frac{\partial g}{\partial z}(t, z) = \frac{1}{2\pi\mathrm{i}} \int_\gamma \frac{g(t, w)}{(w - z)^2}\, dw, \quad z \in D_0. \tag{5.96}$$

As in step 1 and using representation (5.96) we can see that $\partial g/\partial z$ is continuous on $I \times D_0$ and thus on $I \times D$. □

Corollary 5.6.2. *Let D and G be open subsets of $\mathbb{C}$. If γ is a piecewise smooth path in G, and if the continuous function $g : G \times D \to \mathbb{C}$ is analytic with respect to z, then the function f defined by*

$$f(z) := \int_\gamma g(w,z)\,dw, \qquad z \in D$$

is analytic in D and its derivative is given by

$$f'(z) = \int_\gamma \frac{\partial g}{\partial z}(w,z)\,dw.$$

Proof. Assuming that γ is smooth, the result follows on applying Theorem 5.6.1 to the representation

$$f(z) = \int_\gamma g(w,z)\,dw = \int_\alpha^\beta g\big(\gamma(t),w\big)\,\frac{\partial \gamma}{\partial t}\,dt.$$

In the general case, we just add up the integrals along the smooth components of the path. ☐

Example 5.6.1 (Bessel Functions of Integral Order). One source of special functions are eigenvalue problems which arise in mathematical physics. For example, solutions of partial differential equations on circular or cylindrical domains often involve *Bessel functions*. Here we restrict ourselves to Bessel functions of the *first kind* with *integral order* n. These functions can be conveniently defined by the parameter integrals

$$J_n(z) := \frac{1}{2\pi} \int_0^\pi \cos\,(z \sin t - nt)\,dt, \qquad z \in \mathbb{C}. \tag{5.97}$$

Since the integrand is continuous on $[0,\pi] \times \mathbb{C}$ and analytic with respect to z, the assumptions of Theorem 5.6.1 are satisfied, so that the J_n are entire functions if $n \in \mathbb{Z}$. Graphs of the functions J_0 and J_5 are presented in Figure 5.19. Figure 5.20 depicts the corresponding phase portraits in the square $|\mathrm{Re}\, z| < 15$, $|\mathrm{Im}\, z| < 15$.

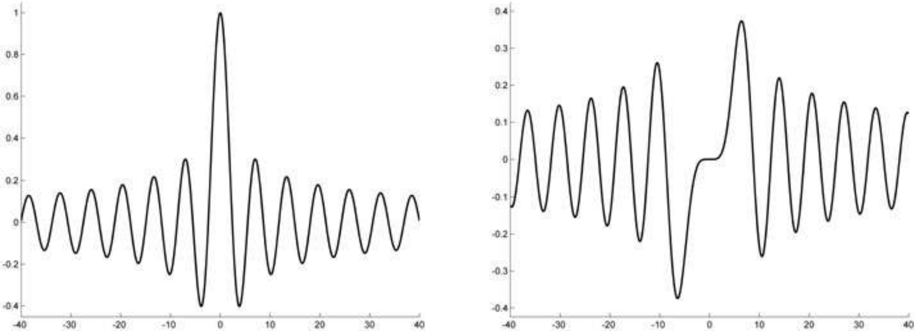


Figure 5.19: Graphs of the real Bessel functions J_0 and J_5

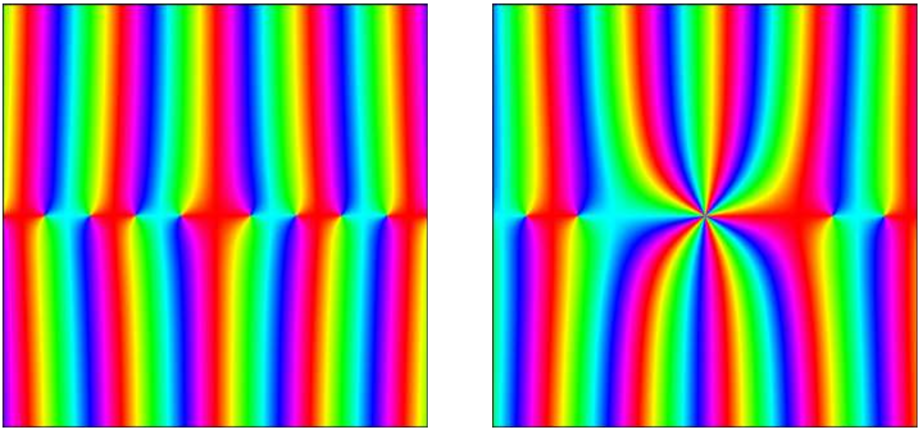


Figure 5.20: Phase portraits of the complex Bessel functions J_0 and J_5

The functions J_n have a zero of order n at the origin and an infinite number of simple zeros along the real axis. The function J_0 looks like a damped cosine function, and in fact the asymptotic behavior of J_n for $x \to +\infty$ is

$$J_n(x) \sim \sqrt{\frac{2}{\pi x}} \cos\left(x - \frac{n\pi}{2} - \frac{\pi}{4}\right).$$

The investigation of Bessel functions and related Neumann and Hankel functions will be continued in Volume 2.

Example 5.6.2 (**The Gamma Function**)**.** One of the most famous functions with a long and interesting history is the *Gamma function.* Originating in the quest for extending the factorial function $f(n) = n\,!$ to positive real arguments, it has been intensely studied by a number of prominent mathematicians. The first satisfactory solution of the factorial problem dates back to Leonhard Euler, who proposed a product representation. Some time later Euler found an integral representation (for real arguments), which is equivalent to the usual definition in present day texts,

$$\Gamma(z) := \int_0^\infty t^{z-1}\, \mathrm{e}^{-t}\, dt, \qquad \operatorname{Re} z > 0. \tag{5.98}$$

Here t^z denotes the principal branch of the complex power function defined by

$$t^z := \exp(z \operatorname{Log} t), \qquad t \in \mathbb{R}_+.$$

The definition of the Gamma function (5.98) involves an *improper integral.* Since $|t^z| = t^{\operatorname{Re} z}$, the integrand decays exponentially as $t \to +\infty$, so that the corresponding integral from 1 to $+\infty$ is always convergent. Near $t = 0$ the situation is a little different. If $\operatorname{Re} z \geq 1$, the integrand is bounded, if $0 < \operatorname{Re} z < 1$ it has an integrable singularity, and if $\operatorname{Re} z \leq 0$ the integral does not converge. So the

Gamma function is well defined by (5.98) in the right half-plane $\operatorname{Re} z > 0$. In order to show that Γ is an analytic function, we consider the truncated integrals

$$\Gamma_n(z) := \int_{1/n}^{n} t^{z-1}\,\mathrm{e}^{-t}\,dt, \qquad \operatorname{Re} z > 0. \tag{5.99}$$

Since the mapping $z \mapsto t^z$ is analytic for any $t \in \mathbb{R}_+$, the functions Γ_n are analytic by Theorem 5.6.1. The standard integral estimate then shows that Γ_n converges normally in the right half-plane to Γ as $n \to \infty$, so that Γ is analytic by Theorem 5.1.3.

So far Γ is only defined on the right half plane. Is there an analytic extension to a larger domain? Analytic continuation by Weierstrass' disk chain method would be technically complicated, so we shall use this opportunity to demonstrate another method. For $0 < a < b < \infty$, *integration by parts* yields

$$\int_a^b t^{z-1}\,\mathrm{e}^{-t}\,dt = \frac{b^z \mathrm{e}^{-b}}{z} - \frac{a^z \mathrm{e}^{-a}}{z} + \frac{1}{z}\int_a^b t^z\,\mathrm{e}^{-t}\,dt. \tag{5.100}$$

If $\operatorname{Re} z > 0$, the boundary terms vanish as $a \to +0$ and $b \to +\infty$, so that we have

$$\Gamma(z) = \int_0^\infty t^{z-1}\,\mathrm{e}^{-t}\,dt = \frac{1}{z}\int_0^\infty t^z\,\mathrm{e}^{-t}\,dt. \tag{5.101}$$

Now the second integral on the right-hand side makes sense for $\operatorname{Re} z > -1$ and defines an analytic function. Consequently (5.101) gives rise to an analytic extension of the function in (5.98) to the domain $\{z \in \mathbb{C} : \operatorname{Re} z > -1\} \setminus \{0\}$. Note that the extended function has a simple pole at 0.

Moreover, the right-hand side of (5.101) is equal to $(1/z)\Gamma(z+1)$, so that we have the *functional equation*

$$\Gamma(z+1) = z\,\Gamma(z). \tag{5.102}$$

Applying this equation recursively in the form $\Gamma(z-1) = \Gamma(z)/(z-1)$, we can extend Γ analytically step by step to the domains

$$\{z \in \mathbb{C} : \operatorname{Re} z > -k\} \setminus \{0, -1, \ldots, -k+1\},$$

and finally to $\mathbb{C} \setminus \{0, -1, \ldots, -k, \ldots\}$. The exceptional points are (simple) poles, so that the resulting Gamma function is meromorphic in the complex plane.

Figure 5.21 shows the graph of the real Gamma function in the interval $[-5, 5]$ and a phase portrait of the complex function in the square $-7 < \operatorname{Re} z < 7$, $-1 < \operatorname{Im} z < 13$. We see the poles at the non-positive integers, but no zeros (in fact the function has no zeros). The function grows (faster than exponentially) along the positive real axis, and decays along the isochromatic lines in the direction of the left upper corner. It is real on the real axis and (therefore) obeys the symmetry relation $\Gamma(\overline{z}) = \overline{\Gamma(z)}$.

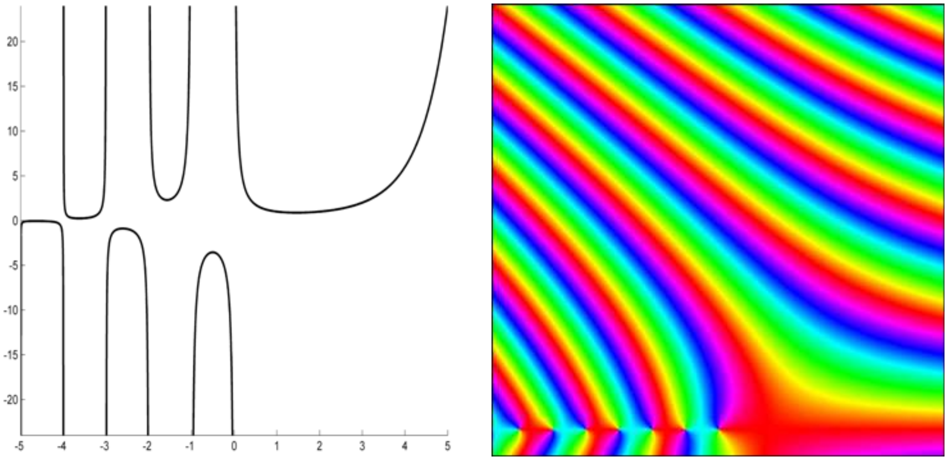


Figure 5.21: Graph on the real axis and phase portrait of the Gamma function

Example 5.6.3 (**The Riemann Zeta Function II**). We continue the investigation of the Riemann Zeta function defined in Example 5.3.3 by a Dirichlet series converging in the half plane $\operatorname{Re} z > 1$. To begin with we derive an (parameter) integral representation which establishes a relation between the Zeta function and the Gamma function, namely we shall show that

$$\zeta(z)\,\Gamma(z) = \int_0^{+\infty} \frac{t^{z-1}}{\mathrm{e}^t - 1}\,dt, \qquad \operatorname{Re} z > 1. \tag{5.103}$$

Here and in the following t^z with $t \in \mathbb{R}_+$ always denotes the main branch of the power function, $t^z := \exp(z \operatorname{Log} t)$. Note that $t^z \to 0$ as $t \to +0$ and $\operatorname{Re} z > 0$.

In order to verify (5.103) we fix some z with $\operatorname{Re} z > 1$. From the definitions (5.22) and (5.99) of ζ and Γ we obtain

$$\begin{aligned} \zeta(z)\,\Gamma(z) &= \sum_{n=1}^{\infty} \frac{1}{n^z} \int_0^{+\infty} t^{z-1}\,\mathrm{e}^{-t}\,dt \\ &= \sum_{n=1}^{\infty} \frac{1}{n} \int_0^{+\infty} \left(\frac{t}{n}\right)^{z-1} \mathrm{e}^{-t}\,dt \\ &= \sum_{n=1}^{\infty} \int_0^{+\infty} t^{z-1}\,\mathrm{e}^{-nt}\,dt. \end{aligned}$$

Assuming, for a moment, that summation and integration can be interchanged, and using the property that the geometric series

$$\sum_{n=1}^{\infty} \mathrm{e}^{-nt} = \frac{1}{\mathrm{e}^t - 1} \tag{5.104}$$

converges uniformly with respect to t on compact subsets of $(0,+\infty)$, we arrive at

$$\zeta(z)\,\Gamma(z)=\int_0^{+\infty} t^{z-1}\sum_{n=1}^{\infty}\mathrm{e}^{-nt}\,dt=\int_0^{+\infty}\frac{t^{z-1}}{\mathrm{e}^t-1}\,dt,$$

which is the desired relation (5.103). In order to justify interchanging the sum and the integral, we consider the absolute values of the summands,

$$\left|t^{z-1}\,\mathrm{e}^{-nt}\right|=t^{\delta}\,\mathrm{e}^{-nt},\qquad \delta:=\operatorname{Re}z-1>0,$$

which implies that for all $N\geq 1$ and $t>0$

$$\sum_{n=1}^{N}\left|t^{z-1}\,\mathrm{e}^{-nt}\right|=\sum_{n=1}^{N}t^{\delta}\,\mathrm{e}^{-nt}\leq\frac{t^{\delta}}{\mathrm{e}^t-1}.$$

Since the function on the right-hand side decays exponentially as $t\to+\infty$ and grows like $t^{\delta-1}$ as $t\to+0$, it is integrable on $[0,+\infty)$. An application of Lebesgue's theorem on dominated convergence completes the proof.

Next we show that the integral on the right-hand side of (5.103) extends to a meromorphic function G in the plane. To this end, we write[3]

$$G(z):=\int_0^{+\infty}\frac{t^{z-1}}{\mathrm{e}^t-1}\,dt=\int_0^{1}\frac{t^{z-1}}{\mathrm{e}^t-1}\,dt+\int_1^{+\infty}\frac{t^{z-1}}{\mathrm{e}^t-1}\,dt=f(z)+g(z).\quad(5.105)$$

The function g defined by the second integral is entire, which follows from a slight modification of Theorem 5.6.1. In order to prove the same for f, we observe that the function h defined by $h(t):=1/(\mathrm{e}^t-1)$ is analytic in the punctured disk $0<|t|<2\pi$ and has a simple pole with residue 1 at the origin. The Laurent series of h has the form

$$h(t)=\frac{1}{\mathrm{e}^t-1}=\frac{1}{t}+\sum_{k=0}^{\infty}a_k t^k,$$

and Cauchy's estimate (Lemma 3.2.4) with $r=2<2\pi$ yields that for some $c>0$

$$|a_k|\leq c\,2^{-k},\qquad k\in\mathbb{N}.\qquad(5.106)$$

It can easily be checked that $h+1/2$ is an odd function, which implies that $a_{2k}=0$, and consequently we have

$$h(t)=\frac{1}{t}-\frac{1}{2}+\sum_{k=0}^{\infty}a_{2k+1}t^{2k+1}.$$

Since z is a number with $\operatorname{Re}z>1$, we have $\left|t^{z+2k}\right|\leq 1$ for $t\in[0,1]$, and using the estimate of the coefficients a_k we see that the series

$$\sum_{k=0}^{\infty}t^{z-1}\,a_{2k+1}t^{2k+1}$$

---

[3] following Ullrich [67], p. 415 f.

converges absolutely and uniformly with respect to t on $[0,1]$. Hence we can exchange integration and summation in the integral defining f,

$$\begin{aligned} f(z) &= \int_0^1 \frac{t^{z-1}}{\mathrm{e}^t - 1}\, dt = \int_0^1 t^{z-2}\, dt - \frac{1}{2}\int_0^1 t^{z-1}\, dt + \sum_{k=0}^{\infty} a_{2k+1} \int_0^1 t^{z+2k}\, dt \\ &= \frac{1}{z-1} - \frac{1}{2z} + \sum_{k=0}^{\infty} \frac{a_{2k+1}}{z+2k+1}. \end{aligned}$$

By Cauchy's estimate (5.106), the series on the right-hand side converges normally to a function which is analytic in $D := \mathbb{C} \setminus \{-2k-1 : k \in \mathbb{N}\}$ and has simple poles at $0, 1$, and possibly (if $a_{2k+1} \neq 0$) at the points $-2k-1$, which proves our claim.

Figure 5.22 (left) shows a phase portrait of G in the square $|\mathrm{Re}\, z| < 20$, $|\mathrm{Im}\, z| < 20$. For comparison, the window on the right depicts the Gamma function in the same domain. Near the right boundary of the squares the portraits are almost identical. Along the real axis we see a number of poles. The poles of G are located at $1, 0, -1, -3, -5, \ldots$, while Γ has poles at $0, -1, -2, -3, \ldots$. In the phase portrait of G we also detect two zeros which are symmetric about the real axis.

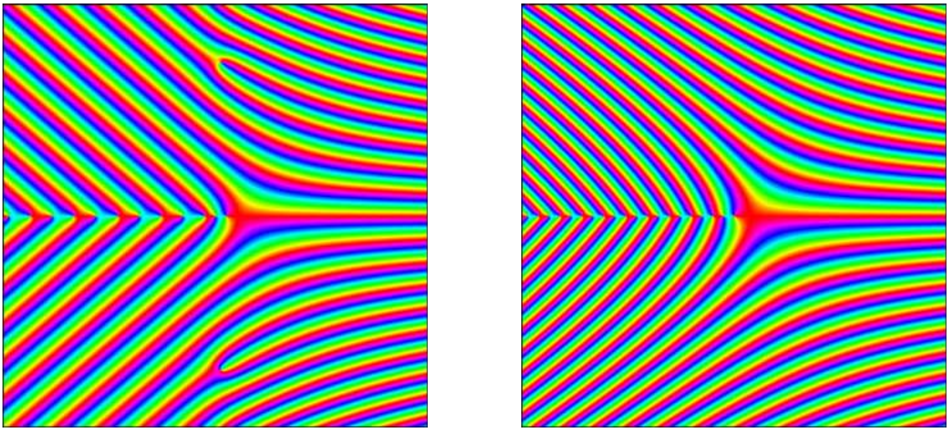


Figure 5.22: Phase portraits of G (left) and Γ (right)

So far we have shown that for all z with $\mathrm{Re}\, z > 1$ the Riemann Zeta function admits the representation

$$\zeta(z) = G(z)/\Gamma(z), \tag{5.107}$$

where G is defined by (5.105) and extends to a meromorphic function in the plane. Since the Gamma function has no zeros, and has simple poles at the non-positive integers, all poles of G except the one at $z_0 = 1$ are cancelled, so that the function G/Γ is meromorphic in $\mathbb{C}$ with a simple pole at 1. Taking (5.107) as a definition yields an analytic continuation of the Riemann Zeta function

to the complex plane. The extended function has simple zeros at the points $-2, -4, -6, \ldots$, which are generated by the poles of Γ.
Figure 5.23 shows the phase portrait of the Riemann Zeta function in the square $|\operatorname{Re} z| < 20$, $|\operatorname{Im} z| < 20$. Besides the so-called "trivial" zeros along the negative real line and the pole at 1 we detect two other zeros (located at $0.5 \pm 14.1347\ldots$ i). Enlarging the domain reveals the existence of many more non-trivial zeros, which seem to be arranged along a straight line. This can be seen more clearly in Figure 5.1 on page 202, which depicts ζ in the region

$$-30 < \operatorname{Re} z < 7, \quad -2 < \operatorname{Im} z < 48.$$

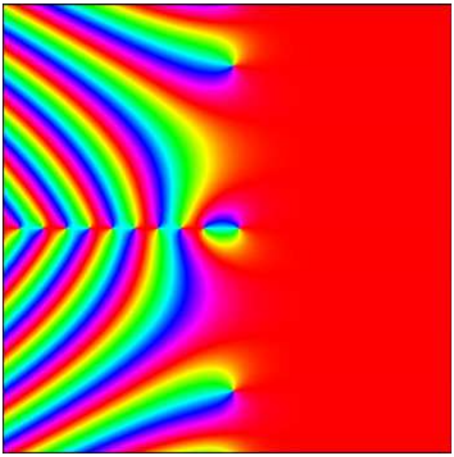


Figure 5.23: The Riemann Zeta function

Since Bernhard Riemann's groundbreaking paper [57] of 1859 it is known that there are indeed infinitely many zeros in the *critical strip*

$$S := \{z \in \mathbb{C} : 0 < \operatorname{Re} z < 1\}$$

and that we find only "trivial" zeros outside the closure of S. Furthermore, Riemann claimed that the number of zeros with $0 < \operatorname{Im} z < T$ should approximately be

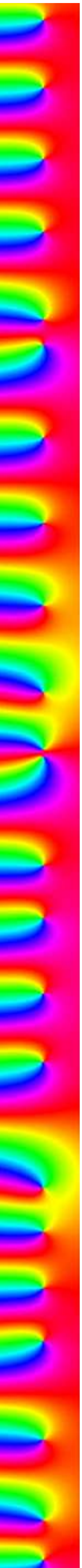
$$\frac{T}{2\pi} \log \frac{T}{2\pi} - \frac{T}{2\pi},$$

but it took almost 50 years until von Mangoldt verified this rigorously. In his paper of less than ten pages Riemann also states that it is "very likely" that all non-trivial zeros lie on the *critical line* $\{z \in \mathbb{C} : \operatorname{Re} z = 1/2\}$. This innocent assertion is the famous *Riemann hypothesis*, which has withstood all the many attempts at proving it, as well as a few at disproving it, for more than 150 years. The beauty and the depth of the Riemann hypothesis combined with the importance it has in almost all fields of mathematics (and even in physics), are some of the reasons why many mathematicians consider it as the holy grail of their science.

The location of non-trivial zeros of the Riemann Zeta function is of great importance in several fields of mathematics, and in particular in number theory. A first hint that the Riemann Zeta function is related to prime numbers comes from a result which was known already to Leonhard Euler, at least for $z \in \mathbb{R}$.

Theorem 5.6.3 (Euler Product Formula)**.** *For all* $z \in \mathbb{C}$ *with* $\operatorname{Re} z > 1$ *we have*

$$\zeta(z) = \prod_{p \in \mathbb{P}} \frac{1}{1 - p^{-z}}, \tag{5.108}$$



where the product extends over the set $\mathbb{P}$ *of all* prime numbers.

Proof. We follow Norton [47]. Multiplying the series (5.22) which defines the Riemann Zeta function by $\big(1-2^{-z}\big)$, we get

$$\zeta(z)(1-2^{-z}) = \sum_{n=1}^{\infty} \frac{1}{n^z} - \sum_{n=1}^{\infty} \frac{1}{(2n)^z} = \frac{1}{1^z} + \frac{1}{3^z} + \frac{1}{5^z} + \frac{1}{7^z} + \dots,$$

where summation on the right-hand side is over the powers of all odd positive integers n. Repeating this with the factor $\big(1-3^{-z}\big)$, we remove in the next step all powers of n which are multiples of 3, so that the series now starts with $1+5^{-z}$. Continuing in this manner, we multiply by $\big(1-m^{-z}\big)$ when the series begins with $1+m^{-z}$, which removes all powers n^{-z} where n is an integral multiple of m.

Because this procedure corresponds exactly to what is done in the sieve of Eratosthenes, it is clear that the numbers m are exactly the prime numbers. Moreover, a simple estimate shows that the sum $1+m^{-z}+\dots$ which remains on the right-hand side tends to 1 as $m\to\infty$.

Since the prime numbers form a subset of the positive integers, and $\sum_{n=1}^{\infty}\big|n^{-z}\big|$ converges for $\operatorname{Re} z > 1$, the product $\prod_{p\in\mathbb{P}}\big(1-p^{-z}\big)$ converges absolutely by Theorem 5.4.4. Finally, the value of this product is the reciprocal of the right-hand side of (5.108). □

The crucial importance of the Riemann Zeta function for the distribution of prime numbers was discovered by Bernhard Riemann in [57]. Riemann verified that all non-trivial zeros of the Zeta function lie in the *closure* of the critical strip, and derived an asymptotic formula for the *prime-counting function* $x\mapsto\pi(x)$, which indicates the number of primes less than or equal to some real number x.

By 1896 the techniques for studying the Zeta function were so advanced that Jacques Hadamard and Charles de la Vallée-Poussin could prove (independently) the *prime number theorem*

$$\pi(x) \sim \frac{x}{\log x},$$

which was already conjectured by Carl Friedrich Gauss. A tough part of these proofs is to show that ζ has no zeros with $\operatorname{Re} z = 1$.

The Riemann Zeta function is of an unbelievable complexity. In order to impart a feeling for the mixture of regularity and (apparent) randomness, which is typical for its behavior, phase portraits of Zeta in the critical strip are shown in the margins of several pages of this book. [4] The range depicted on page p is approximately the rectangle

$$\{z\in\mathbb{C} : 0 < \operatorname{Re} z < 1,\ 20\,p-21 < \operatorname{Im} z < 20\,p+1\}.$$

---

[4] You need not search for zeros violating the Riemann hypothesis, it has been verified that there is no counterexample among the first 10,000,000,000,000 zeros, at least.

The Zeta function will be investigated further in Volume 2. In particular we shall explain why it may be considered as the *mother* of all (analytic) functions.

Readers who want to know more about the history of and contemporary developments in this fascinating field will find quite a number of excellent and entertaining books, like Derbyshire [12], Sabbagh [59] and du Sautoy [61]. A more advanced introduction is given by Borwein, Choi, Rooney and Weirathmueller [5], including reprints of many original papers which are milestones in the exploration of the Riemann Zeta function.

Integral Transforms. We cannot finish this section without mentioning that parameter integrals are the basis of various *integral transforms*, like the *Laplace*, the *Fourier*, and the *Mellin transform*, which are defined by the parameter integrals

$$\int_0^\infty f(t)\,\mathrm{e}^{-tz}\,dt, \qquad \int_{-\infty}^\infty f(t)\,\mathrm{e}^{\mathrm{i}tz}\,dt, \qquad \int_0^\infty f(t)\,t^{z-1}\,dt,$$

respectively. Several examples of such integrals already came up in the text: in Example 4.5.5 we have calculated the Fourier transform $F(z) = \pi\mathrm{e}^{-|z|}$ of the function $f(t) = 1/(1+t^2)$ for real arguments z, the Gamma function Γ is by definition the Mellin transform of $f(t) = \mathrm{e}^{-t}$, and formula (5.103) tells us that its product $\zeta\,\Gamma$ with the Riemann Zeta function is the Mellin transform of the function $f(t) = 1/(\mathrm{e}^{-t}-1)$.

Properties of integral transforms and some of their applications will be discussed in more detail in Volume 2.

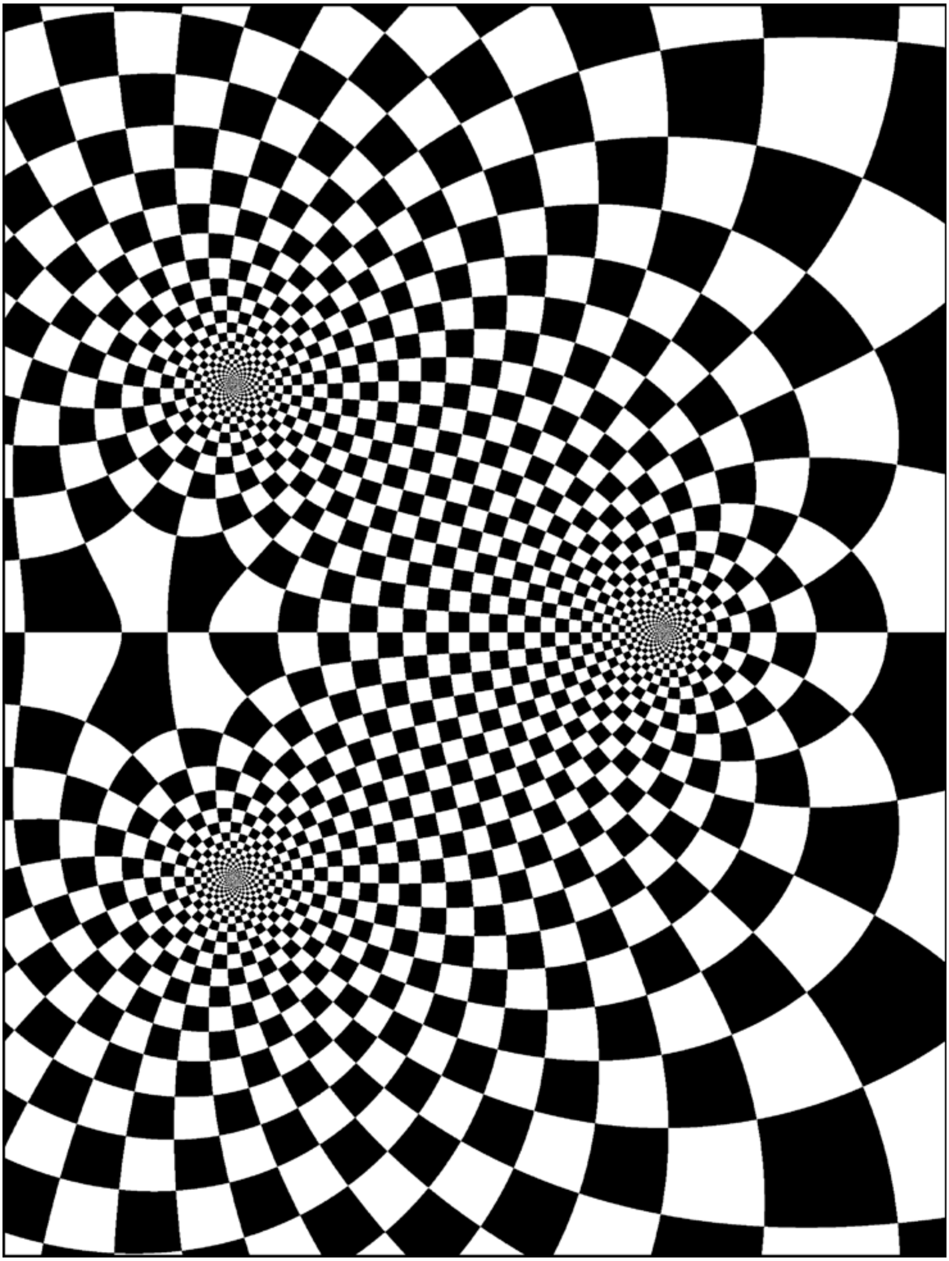


Figure 6.1: Pull-back of a polar chessboard via $f(z) = (z-1)/(z^2+z+1)$

Chapter 6

Conformal Mappings

The focus of this chapter is on geometric aspects: complex functions will be considered as mappings between subsets of two complex planes which transplant objects from one plane to the other.

In the first section we show that analytic functions are characterized by the property of *conformality* – which means that the associated mappings preserve the angles of intersection between smooth curves.

In Section 6.2 we study analytic functions which map two given domains conformally onto each other. We describe the conformal self-mappings of the sphere, the plane and the disk, and we search through the inventory of elementary functions for appropriate mappings between some standard domains.

Section 6.3 is exclusively devoted to Möbius transformations, the conformal automorphisms of the Riemann sphere. We introduce and study the cross ratio, and classify Möbius transformations according to their invariant sets.

In Section 6.4 we prove the celebrated Riemann mapping theorem on the existence of conformal mappings between arbitrary simply connected domains. We present an elegant proof published by Constantin Carathéodory in 1928, more than 75 years after the result was first stated by Bernhard Riemann.

A second theorem of almost the same caliber is the Carathéodory–Osgood theorem on the boundary correspondence of conformal mappings between Jordan domains. We state and prove this result in Section 6.5.

Section 6.6 is devoted to the Schwarz reflection principle, another technique for extending analytic functions to larger domains, and its applications to conformal mappings between symmetric domains. A more serious usage of this principle follows in Section 6.7, where we study the mapping properties of elliptic integrals and their inverses, which turn out to be doubly periodic analytic functions. In the last section these ideas are put into a more general context, leading to the Schwarz–Christoffel formula for conformal mappings of a half-plane onto polygonal domains.

E. Wegert, *Visual Complex Functions: An Introduction with Phase Portraits*,
DOI 10.1007/978-3-0348-0180-5_6, © Springer Basel 2012

6.1 Mappings of Planar Domains

In this section we investigate why (almost all) ‘tiles’ of the phase portraits with enhanced contour lines of modulus and phase of analytic functions f are approximately square. Figure 6.2 shows a portrait of the standard example from Section 2.5, $f(z) := (z-1)/(z^2+z+1)$. In order to focus on the shape of the tiles, the image plane in the picture on the right is not colored according to phase, but so that all tiles are colored in the same manner. Four different colors at the corners allow one to identify the orientation of the tiles. In the picture on the left, this coloring is pulled back to the domain of f.

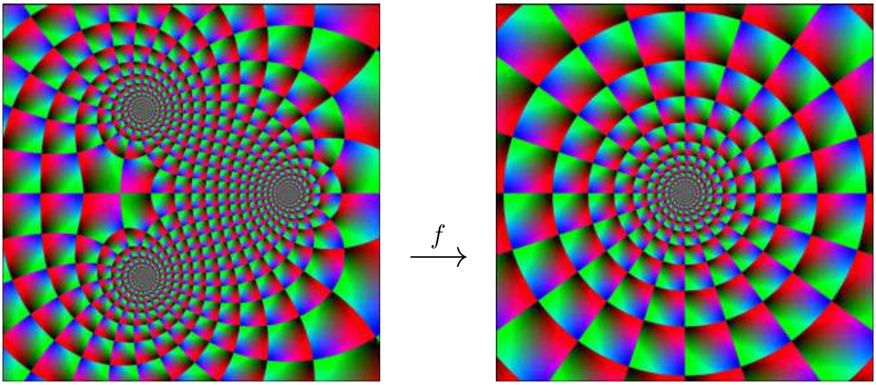


Figure 6.2: A rectangular tiling of the image plane pulled back to the domain

A rough explanation for the shape of the transplanted tiles has already been given in Section 4.1: an analytic function with $f'(z_0) \neq 0$ behaves locally at z_0 like the linear mapping $z \mapsto f(z_0) + f'(z_0)\,(z - z_0)$, which is an orientation preserving similarity transformation. However, this does not explain why the corners of the tiles are *exactly* rectangular.

Preservation of Angles. In order to study this in more detail, we consider the transplantation of smooth (oriented) curves from the z-plane to the w-plane under a general complex mapping $f : D \to \mathbb{C}$. In the following we make the standing assumption that $f : D \to \mathbb{C}$ is an $\mathbb{R}$-differentiable function (see Definition 4.1.9). Since it requires a little less machinery, we shall work with paths instead of curves.

To begin with we fix a point $z \in D$ and consider a differentiable path $\gamma : [0,1] \to D$ with initial point z. We may assume that γ is simple, though this is not really important. If $\gamma'(0) \neq 0$, then the vector $v := \gamma'(0)$ is tangent to the trace of γ at z. Regardless of whether $v = 0$ or $v \neq 0$ we call v the tangent vector of γ at z.

Let now γ_1 and γ_2 be two differentiable paths in D with initial point z and assume that the tangent vectors $v_1 := \gamma_1'(0)$ and $v_2 := \gamma_2'(0)$ at z are different from zero. Then the *angle of intersection* of γ_1 and γ_2 at z is defined as the *oriented* angle between the vectors v_1 and v_2,

$$\alpha = \angle(\gamma_1, \gamma_2) := \operatorname{Arg}\big(v_2/v_1\big) \in (-\pi, \pi].$$

Next we compare the intersection angle α of γ_1 and γ_2 at z with the intersection angle $\beta = \angle\big(f \circ \gamma_1, f \circ \gamma_2\big)$ of the transplanted paths $f \circ \gamma_1$ and $f \circ \gamma_2$ at $f(z)$. This is illustrated in Figure 6.3.

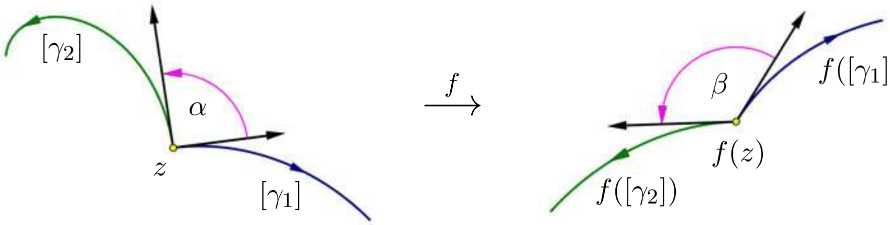


Figure 6.3: Intersection angles of two differentiable paths and their images

In the case depicted the angles α and β are different. Mappings for which α and β are always equal are of special interest.

Definition 6.1.1. Let γ_1 and γ_2 be differentiable paths in D with initial point z and non-vanishing tangent vectors at z. We say that a mapping $f : D \to \mathbb{C}$ preserves the angle of intersection of γ_1 and γ_2 if the tangent vectors of $f \circ \gamma_1$ and $f \circ \gamma_2$ at $f(z)$ are non-zero and

$$\angle(f \circ \gamma_1, f \circ \gamma_2) = \angle(\gamma_1, \gamma_2).$$

The mapping f is said to be *conformal* at z, if it preserves the angles of intersection of arbitrary differentiable paths with non-vanishing tangent vectors at z. A function which is conformal at every point in D is called *conformal in* D.

Note that, according to this definition, conformal mappings preserve *oriented* angles. For example, reflections along a line are *not conformal* because they reverse orientation. A more general concept neglects orientation and requires only that $\big|\angle(f \circ \gamma_1, f \circ \gamma_2)\big| = \big|\angle(\gamma_1, \gamma_2)\big|$. We then get two classes of angle preserving mappings, which are related to each other by reflection along a straight line, for example the real axis.

The next theorem provides a general criterion for conformality which relates it to a familiar class of functions.

Theorem 6.1.2. *Let $f : D \to \mathbb{C}$ be an $\mathbb{R}$-differentiable function in an open set $D \subset \mathbb{C}$. Then f is conformal at z if and only if f is complex differentiable at z and $f'(z) \neq 0$. The mapping f is conformal in D if and only if f is analytic in D and $f'(z) \neq 0$ for all $z \in D$.*

Proof. Let $v = \xi + \mathrm{i}\eta$ be the tangent vector of a differentiable path γ at $z := \gamma(0)$. By the chain rule, the tangent vector of $f \circ \gamma$ at $f(z)$ is given by

$$(f \circ \gamma)'(0) = (a\xi + b\eta) + \mathrm{i}(c\xi + d\eta), \tag{6.1}$$

where the coefficients a, b, c, d are the entries of the *Jacobi matrix*

$$Df(z) := \begin{bmatrix} \dfrac{\partial u}{\partial x}(x,y) & \dfrac{\partial u}{\partial y}(x,y) \\ \dfrac{\partial v}{\partial x}(x,y) & \dfrac{\partial v}{\partial y}(x,y) \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \tag{6.2}$$

of f at $z = x + \mathrm{i}y$. Consequently f is conformal at z if and only if the (real) *linear* mapping

$$\begin{bmatrix} \xi \\ \eta \end{bmatrix} \mapsto \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} \xi \\ \eta \end{bmatrix}$$

preserves the oriented angles between vectors in $\mathbb{R}^2$. It is well known (and can be easily verified) that this happens if and only if $Df(z)$ is an orthogonal matrix with positive determinant, i.e., $a = d$, $b = -c$, and $a^2 + b^2 > 0$.

The first two conditions are just the Cauchy-Riemann equations for f at z, which, by Theorem 4.1.10, are equivalent to the complex differentiability of f at z. Expressing f' by the partial derivatives of u we get

$$|f'(z)|^2 = \left(\frac{\partial u}{\partial x}(z)\right)^2 + \left(\frac{\partial u}{\partial y}(z)\right)^2 = a^2 + b^2, \tag{6.3}$$

so that the third condition $a^2 + b^2 > 0$ is equivalent to $f'(z) \neq 0$. ☐

Since all regular parameterizations of a smooth oriented curve have parallel tangent vectors, Theorem 6.1.2 can be rephrased more geometrically as a statement about the preservation of intersection angles between smooth oriented curves.

Remark 6.1.3. The Jacobian J_f, i.e., the determinant of the Jacobi matrix Df, of an $\mathbb{R}$-differentiable function f can be conveniently expressed by the Wirtinger derivatives (4.11),

$$J_f = \left(\frac{\partial f}{\partial z}\right)^2 - \left(\frac{\partial f}{\partial \overline{z}}\right)^2. \tag{6.4}$$

If f is analytic, then $J_f = |f'|^2$. Note that $|J_f(z)|$ is the local scaling factor by which the mapping f expands or shrinks area measure at z.

Example 6.1.1 (The Square Function). Figure 6.4 illustrates how the function $f(z) = z^2$ acts on a square grid consisting of two families of horizontal and vertical lines in the domain $0 < \operatorname{Re} z < 1$, $0 < \operatorname{Im} z < 1$. In the region depicted, the function is injective and its derivative does not vanish.

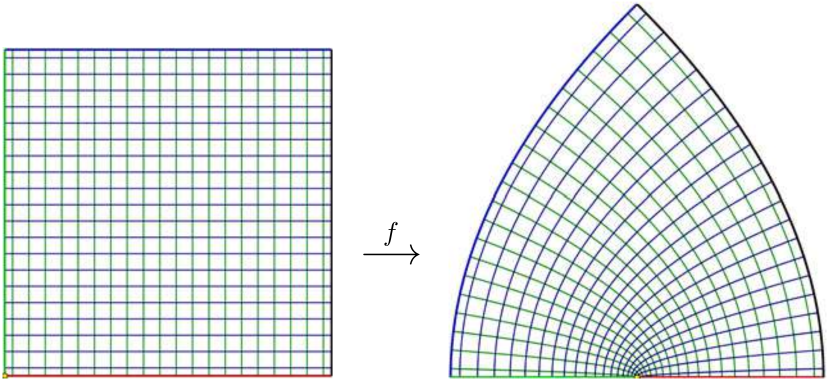


Figure 6.4: A square mesh and its image under the mapping $f(z) = z^2$

The mapping $f(z) = z^2 = x^2 - y^2 + 2\mathrm{i}\,xy$ is conformal at all points of the plane, except at the origin, where its derivative f' has a zero of first order. Figure 6.4 shows that the right angle between the red and the green curve at $z = 0$ is doubled in the image.

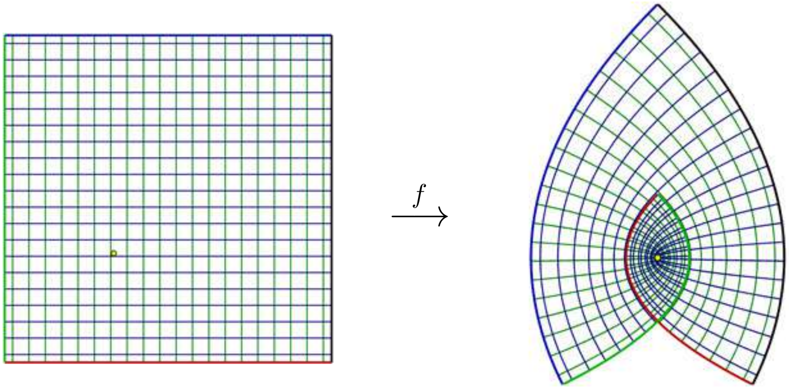


Figure 6.5: Pushing forward a rectangular mesh by $f(z) = z^2$

Critical Points. The only points where an analytic function $f : D \to \mathbb{C}$ is not conformal are the zeros of its derivative.

Definition 6.1.4. The zeros of the derivative f' of an analytic function f are said to be the *critical points* of f. The values of f at its critical points are called *critical values*.

Figure 6.5 illustrates the action of $f(z) = z^2$ on a rectangular grid in some square which contains the critical point $z = 0$ (indicated by the yellow dot). The function f is not injective and the right picture shows clearly that the image domain ‘wraps twice around’ the point $f(0)$. If f' has a zero of order k at z_0, it follows from the representation $f(z) = f(z_0) + (z - z_0)^{k+1} g(z)$ with $g(z_0) \neq 0$ that angles α between curves intersecting each other at z_0 are multiplied by a factor of $k+1$. If $(k+1)\alpha > 2\pi$, this angle has to be measured between the image curves on the corresponding Riemann surface (see Chapter 7), when the curves are considered in the plane as usual, the value of the angle is reduced modulo $2\,\pi$.

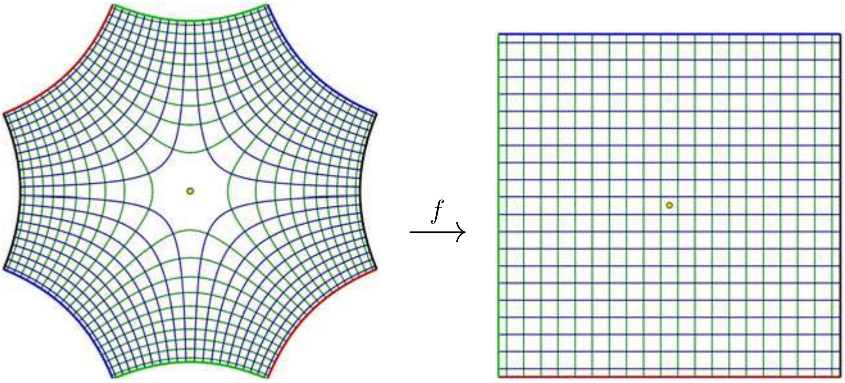


Figure 6.6: Pulling back a rectangular mesh via $f(z) = z^2$

Pushing Forward and Pulling Back. The self-overlapping of images makes “pushing forward” geometric objects not very convenient, while “pulling back” avoids this drawback. This can be seen in Figure 6.6 depicting the pull-back of a mesh of equidistant lines via the mapping $f(z) = z^2$.

The preimages of these lines are the contour lines of the real part u and the imaginary part v of f. Since the lines $\operatorname{Re} w = \text{const}$ and $\operatorname{Im} w = \text{const}$ are mutually orthogonal in the w-plane, we obtain again that the contour lines of conjugate harmonic functions u and v are orthogonal (see Theorem 4.6.3).

Simililarly, the preimage of an *orthogonal polar grid* consists of two families of curves which are mutually orthogonal. Exceptions are not only the critical points of f, but also the zeros of f, which are sent to the center of the polar grid. This again explains the orthogonality of the contour lines of phase and modulus in the phase portrait of analytic functions (see the discussion at the end of Section 4.1).

Example 6.1.2. As an example of a mapping which is not conformal, we consider the function $g(x + \mathrm{i}\, y) := x^2 - y^2 + \mathrm{i}\,(x + y)$. It is not analytic and its Jacobian

$$Dg(x + \mathrm{i}\, y) = \begin{bmatrix} 2x & -2y \\ 1 & 1 \end{bmatrix}$$

is an orthogonal matrix only at the point $z = (1+\mathrm{i})/2$. Figure 6.7 shows the action of g on a mesh in the square $0 < \operatorname{Re} z < 1$, $0 < \operatorname{Im} z < 1$. The red cells contain the point $z = (1+\mathrm{i})/2$ (left), and its image (right). Both are (almost) squares. Note the differences in the shapes of the other three colored cells.

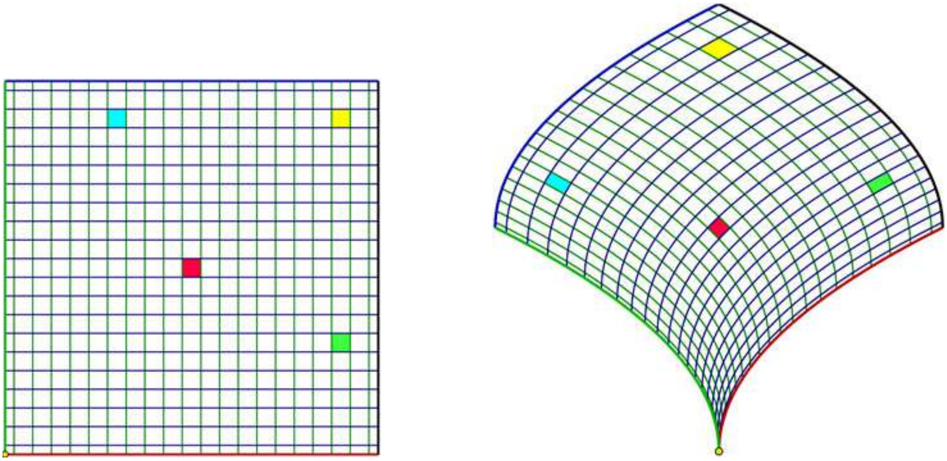


Figure 6.7: A square mesh and its image under the mapping g

6.2 Special Conformal Mappings

We begin by extending the definition of conformality to include the point at infinity in the domain and/or the range of f. Since the (reflected) inversion $z \mapsto 1/z$ is angle preserving on the Riemann sphere, it is natural to say that f is conformal at ∞ if $z \mapsto f(1/z)$ is conformal at 0. Similarly, f is conformal at z_0 with $f(z_0) = \infty$ if $z \mapsto 1/f(z)$ is conformal at z_0.

Using Theorem 6.1.2 it can easily be verified that $f : D \to \widehat{\mathbb{C}}$ is conformal at a point $z_0 \in D \subset \widehat{\mathbb{C}}$ with $w_0 := f(z_0) \in \widehat{\mathbb{C}}$ if and only if it is analytic at z_0 and has the Laurent expansion

$$\begin{array}{llll}
f(z) = c_0 + c_1(z - z_0) + \ldots & \text{with } c_1 \neq 0 & \text{if} & z_0 \in \mathbb{C},\ \ w_0 \in \mathbb{C}, \\
f(z) = \dfrac{c_{-1}}{z - z_0} + c_0 + c_1(z - z_0) + \ldots & \text{with } c_{-1} \neq 0 & \text{if} & z_0 \in \mathbb{C},\ \ w_0 = \infty, \\
f(z) = c_0 + \dfrac{c_{-1}}{z} + \dfrac{c_{-2}}{z^2} + \ldots & \text{with } c_{-1} \neq 0 & \text{if} & z_0 = \infty,\ w_0 \in \mathbb{C}, \\
f(z) = c_1\, z + c_0 + \dfrac{c_{-1}}{z} + \ldots & \text{with } c_1 \neq 0 & \text{if} & z_0 = \infty,\ w_0 = \infty.
\end{array}$$

Univalent Conformal Mappings. So far our definitions of conformality do not require that the mapping be injective. Conformal mappings with this additional property are said to be *univalent* and deserve special attention. In this case the function $f : D \to G$ is a bijective mapping of D onto the image set $G := f(D)$, which is in fact a domain by the open mapping principle (Theorem 3.4.8). Since the converse is true as well, the univalent conformal mappings of D onto G are exactly the bijective analytic functions $f : D \to G$.

Definition 6.2.1. An analytic function $f : D \to G$ which maps a domain D bijectively onto (a domain) G is called a *conformal mapping of D onto G*. If such a mapping exists, G and D are said to be *conformally equivalent*. A *conformal automorphism* of D is a conformal mapping of D onto itself.

Conformal Automorphisms. Because the inverse function $f^{-1} : G \to D$ of a bijective analytic function $f : D \to G$ is automatically continuous, it follows that conformal equivalence is indeed an equivalence relation. The conformal automorphisms of a domain form a group with respect to composition.

Theorem 6.2.2. *Let D be one of the following domains: the Riemann sphere, the complex plane, or the unit disk. Then the conformal automorphisms of D are precisely the following functions:*

(i) *the Möbius transformations $(az+b)/(cz+d)$ with $ad \neq bc$, if $D = \widehat{\mathbb{C}}$;*

(ii) *the linear functions $az+b$ with $a \neq 0$, if $D = \mathbb{C}$;*

(iii) *the Blaschke factors $c\,(z-a)/(1-\overline{a}z)$ with $|a| < 1$ and $|c| = 1$, if $D = \mathbb{D}$.*

Proof. 1. Let $D = \widehat{\mathbb{C}}$. It has already been shown in Section 2.3, that any Möbius transformation maps $\widehat{\mathbb{C}}$ bijectively onto itself. Conversely, any conformal automorphism f of $\widehat{\mathbb{C}}$ must be a rational function (Theorem 3.5.8). Because f is bijective, it must have degree 1 (Theorem 3.1.10), i.e., $f(z) = (az+b)/(cz+d)$ with co-prime linear polynomials $az+b$ and $cz+d$. This also implies that $ad \neq bc$.

2. It is clear that any linear function $f(z) = az+b$ with $a \neq 0$ is a conformal automorphism of $\mathbb{C}$. Conversely, if f is such an automorphism, it has exactly one zero $z_0 \in \mathbb{C}$. So f can be represented as $f(z) = (z-z_0)\,g(z)$, where g is analytic in $\mathbb{C}$ and $g(z_0) \neq 0$. Consequently there exists some neighborhood U of z_0 where $1/g$ is bounded.

By the open mapping principle $f(U)$ covers a neighborhood V of 0, and since the sets $f(U)$ and $f(\mathbb{C} \setminus U)$ are disjoint, $1/f$ is bounded on $\mathbb{C} \setminus U$. This implies that $1/g$ is also bounded on $\mathbb{C} \setminus U$ and hence on all of $\mathbb{C}$. Liouville's theorem (Theorem 4.2.25) tells us that g is constant, i.e., $f(z) = a(z-z_0)$. It is clear that a cannot be zero.

3. The assertion for the unit disk has already been verified in Example 3.4.4 of Section 3.4. □

Conformal Automorphisms of the Unit Disk. In the first and the third cases of the last theorem something more can be said. Here we consider only the case of the disk, where the conformal automorphisms are Blaschke factors,

$$f(z) = c\,\frac{z-z_0}{1-\overline{z}_0\,z}, \qquad |c| = 1,\ |z_0| < 1. \tag{6.5}$$

Möbius transformations will be investigated in detail in the next section.

Theorem 6.2.3. *Let* $z_0 \in \mathbb{D}$ *and* $\alpha \in \mathbb{R}$. *Then there is exactly one conformal automorphism of* $\mathbb{D}$ *with*

$$f(z_0) = 0, \qquad \arg f'(z_0) = \alpha, \tag{6.6}$$

namely the function f *of* (6.5) *with* $c := \mathrm{e}^{\mathrm{i}\alpha}$.

Proof. Any conformal automorphism of $\mathbb{D}$ with $f(z_0) = 0$ has the form (6.5). Since $f'(z_0) = c/\big(1 - |z_0|^2\big)$ and $|c| = 1$, the condition $\arg f'(z_0) = \alpha$ holds if and only if $c = \mathrm{e}^{\mathrm{i}\alpha}$. ☐

Corollary 6.2.4. *If* $z_0 \in \mathbb{D}$ *and* f *is a conformal automorphism of* $\mathbb{D}$ *with* $f(z_0) = z_0$ *and* $f'(z_0) > 0$ *then* $f(z) = z$ *for all* $z \in \mathbb{D}$.

Proof. Let g be a conformal automorphism of $\mathbb{D}$ with $g(z_0) = 0$ and $g'(z_0) > 0$. Since $h := g \circ f$ also satisfies $h(z_0) = 0$ and $h'(z_0) > 0$, we have $h = g$. ☐

Figure 6.8 visualizes a conformal automorphism of the unit disk. The point z_0 and the origin are marked by a yellow dot.

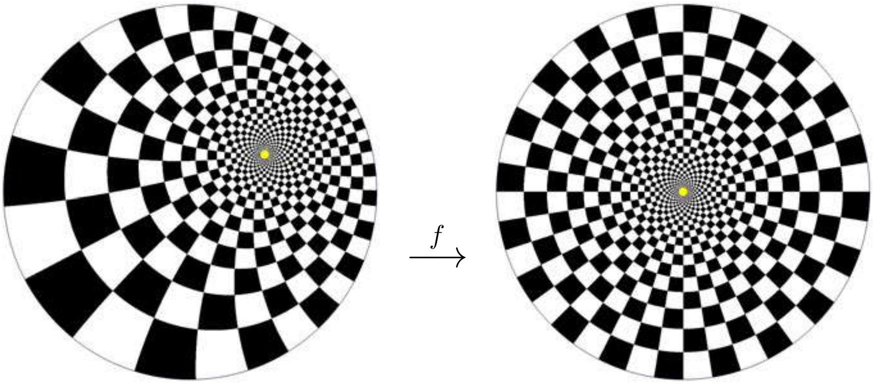


Figure 6.8: A conformal automorphism of the unit disk $\mathbb{D}$

One fundamental question in the theory of conformal mappings is to decide which domains are conformally equivalent. Before we treat this problem in a general setting in Section 6.4, let us develop some useful techniques for constructing special conformal mappings. To start, let us browse through our repertoire of analytic functions for something appropriate.

Example 6.2.1 (The Cayley Mapping). A conformal mapping of the upper half-plane $\mathbb{H}$ onto the unit disk $\mathbb{D}$ is given by the *Cayley mapping*,

$$f(z) = \frac{z - \mathrm{i}}{z + \mathrm{i}}. \tag{6.7}$$

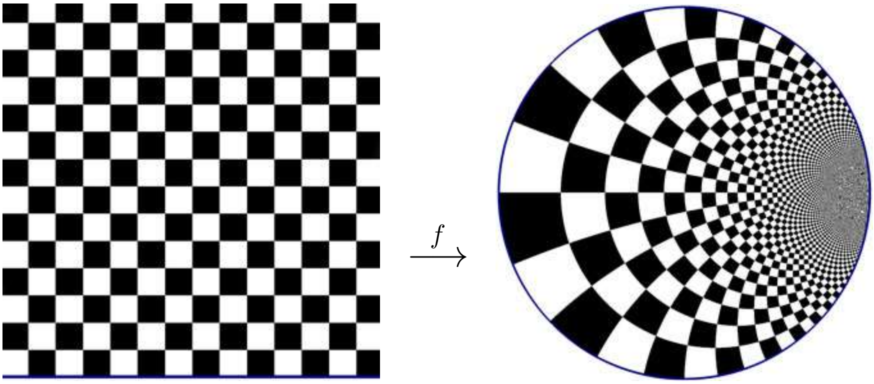


Figure 6.9: The Cayley mapping of a semi-infinite cartesian chessboard

It sends i to 0 and satisfies $f'(\mathrm{i}) = -\mathrm{i}/2$. By Theorem 6.2.3, any conformal mapping of $\mathbb{H}$ onto $\mathbb{D}$ is the post-composition of the Cayley mapping with a Blaschke factor.

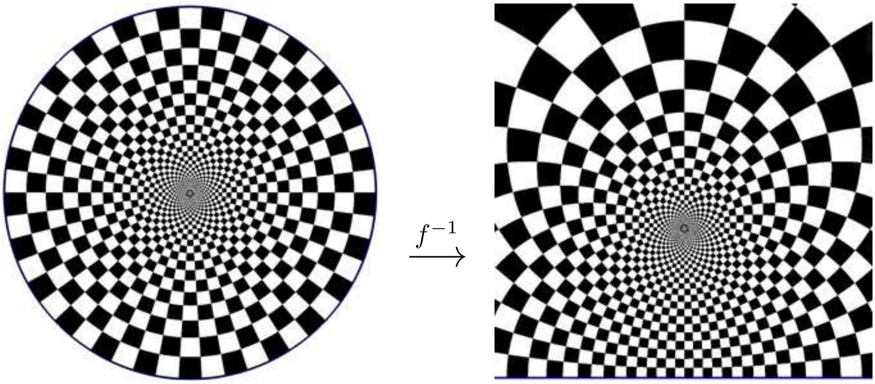


Figure 6.10: The inverse Cayley mapping of a polar chessboard

Figure 6.10 visualizes the inverse of the Cayley mapping f, which is given by

$$f^{-1}(w) = \mathrm{i}\,\frac{1+w}{1-w}. \tag{6.8}$$

Example 6.2.2 (Exponential and Logarithm Function). The *exponential function* and the *logarithm* generate several interesting conformal mappings. It follows from the representation

$$\mathrm{e}^{x+\mathrm{i}y} = \mathrm{e}^{x}\,(\cos y + \mathrm{i}\sin y)$$

that, for all $a, b, c, d \in \mathbb{R}$ with $a < b$ and $-\pi \le c < d \le \pi$, the function $w = \mathrm{e}^{z}$ maps the rectangle

$$R(a, b, c, d) := \{z \in \mathbb{C} : a < \operatorname{Re} z < b,\ c < \operatorname{Im} z < d\}, \tag{6.9}$$

conformally onto the annular sector

$$S(a,b,c,d) := \{w \in \mathbb{C} \setminus \{0\} : \mathrm{e}^a < |w| < \mathrm{e}^b,\ c < \operatorname{Arg} w < d\}. \tag{6.10}$$

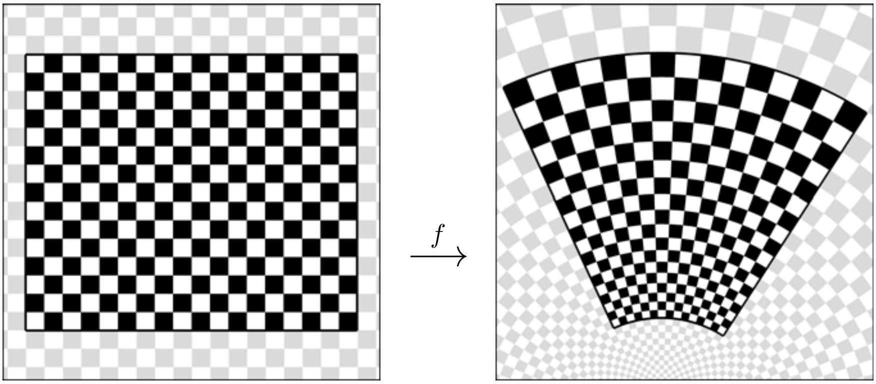


Figure 6.11: The exponential function mapping a rectangle to an annular sector

The infinite strip $R(-\infty,\infty,c,d)$ is mapped by the exponential function onto the sector $\{w \in \mathbb{C} : c < \operatorname{Arg} w < d\}$. In particular, the strip $R(-\infty,\infty,-\pi,\pi)$ is mapped onto the complex plane slit along the negative real axis.

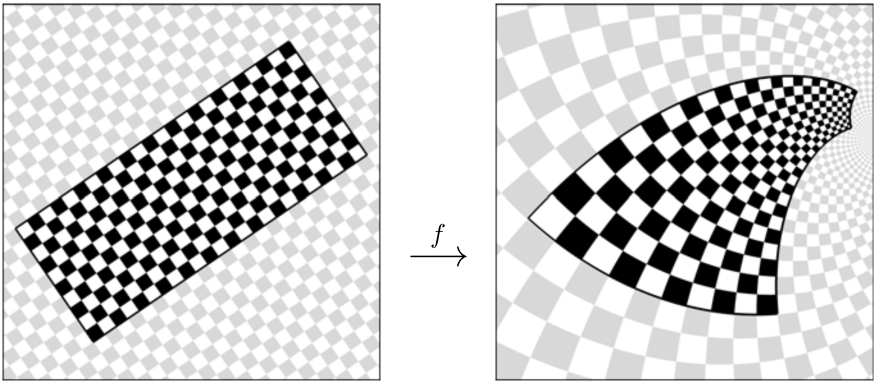


Figure 6.12: The exponential function mapping a rectangle to a spiral domain

The image of a straight line, which is neither parallel to the real nor to the imaginary axis, under the mapping $z \mapsto \mathrm{e}^z$ is a *logarithmic spiral*. Thus the exponential function generally maps a rectangle to a curvilinear quadrilateral bounded by logarithmic spirals. This mapping is univalent if and only if any intersection of the rectangle with a vertical line has length not exceeding 2π. The corresponding inverse mappings are given by appropriately chosen branches of the logarithm.

Example 6.2.3 (Power Functions). A branch of the *general power function* z^α with complex exponent a is defined by

$$f(z) = z^a := \exp(a\,\log z)$$

where log denotes an analytic branch of the logarithm. In order to understand the mapping properties of $z \mapsto w = f(z)$ we decompose it into the functions

$$\begin{aligned} z &\mapsto w_1 := \log z, \\ w_1 &\mapsto w_2 := a\,w_1, \\ w_2 &\mapsto w \;\; := \exp w_2. \end{aligned} \tag{6.11}$$

The sequence of pictures on the right-hand side shows the consecutive images of some sector with vertex at the origin in the z, w_1, w_2, and w-plane, respectively. First, the logarithm maps that sector onto a strip parallel to the real axis, then multiplication by a effects a roto-stretch, and finally the exponential function sends the strip to a domain bounded by two logarithmic spirals.
If a is positive, $-\pi < \alpha < \beta < \pi$ and $-\pi < a\alpha < a\beta < \pi$, the sector

$$\{z \in \mathbb{C} : \;\alpha < \operatorname{Arg} z < \beta\},$$

is mapped by the principal branch, defined by $z^a := \exp(a\operatorname{Log} z)$, onto the sector

$$\{w \in \mathbb{C} : \;a\,\alpha < \operatorname{Arg} w < a\,\beta\}.$$

The mapping properties of the other branches can be investigated analogously. Note that in all cases where a is not an integer, the power functions $f(z) = z^a$ can be studied more conveniently and more naturally on Riemann surfaces (see Chapter 7, in particular Example 7.3.2).

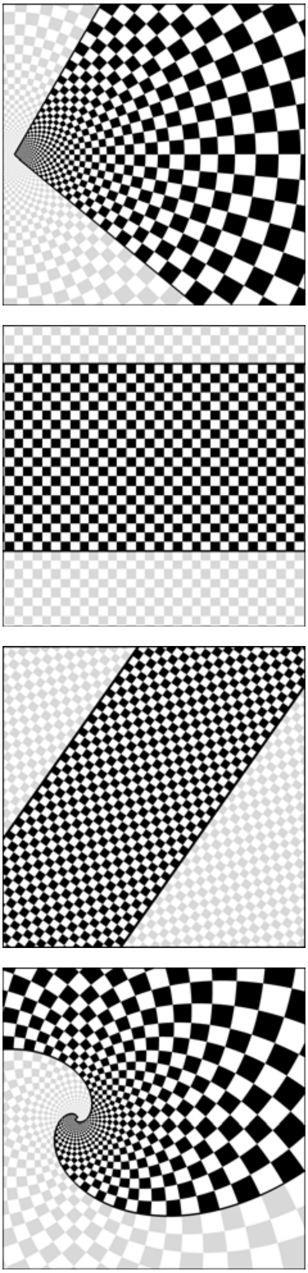


Figure 6.13: A power function

Example 6.2.4 (The Sine Function). In order to find some interesting conformal mappings involving trigonometric functions, we look at the phase portrait of the sine function. Figure 6.14 suggests that the half-strip

$$H := \{x + \mathrm{i}\, y \in \mathbb{C} : \ -\pi/2 < x < \pi/2, \ \ y > 0\}$$

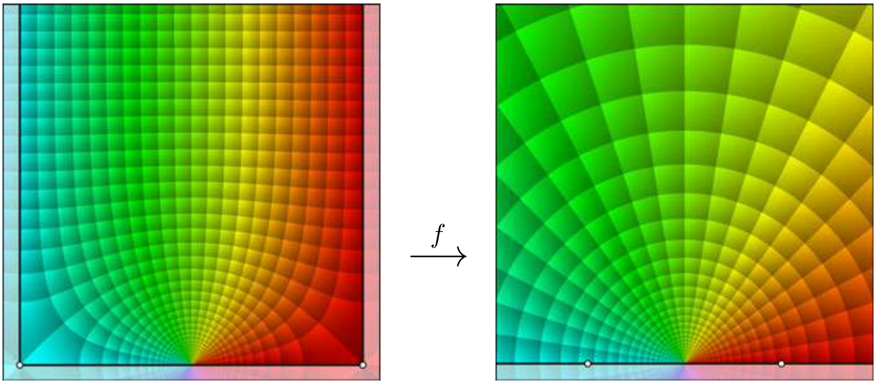


Figure 6.14: Phase portrait and mapping properties of the sine function

is mapped onto the upper half-plane $\mathbb{H}$. This can be confirmed easily using the representation

$$\sin(x + \mathrm{i}\, y) = \sin x \cos(\mathrm{i}y) + \cos x \sin(\mathrm{i}y) = \frac{\mathrm{e}^y + \mathrm{e}^{-y}}{2}\, \sin x + \mathrm{i}\, \frac{\mathrm{e}^y - \mathrm{e}^{-y}}{2}\, \cos x.$$

It tells us that the horizontal segments with $-\pi/2 < x < \pi/2$, $y = \text{const} = d > 0$ are mapped onto semi-ellipses with foci ± 1 and semi-axis $\frac{1}{2}\,(\mathrm{e}^d + \mathrm{e}^{-d})$, $\frac{1}{2}\,(\mathrm{e}^d - \mathrm{e}^{-d})$. The family of these semi-ellipses covers the upper half-plane $\mathbb{H}$.

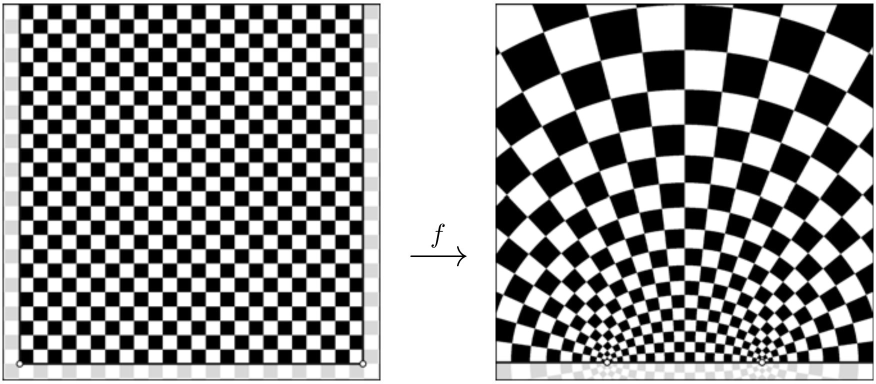


Figure 6.15: The sine function mapping a half-strip onto the upper half-plane

Analogously, the family of vertical half-lines $x = \text{const} = c \in (-\pi/2, \pi/2)$, $y > 0$ is mapped onto a family of semi-hyperbolas covering $\mathbb{H}$. Both families can be easily detected in Figure 6.15, which visualizes the conformal mapping $z \mapsto \sin z$ by transplanting a chessboard from the half-strip to the upper half-plane.

Example 6.2.5 (The Tangent Function). Figure 3.20 on page 90 shows a phase portrait of the complex *tangent function*. The vertical strip $-\pi/2 < \operatorname{Re} z \leq \pi/2$ is mapped bijectively onto the Riemann sphere without the points i and $-$i. The strip $-\pi/4 < \operatorname{Re} z < \pi/4$ is mapped conformally onto the unit disk, as is illustrated by the phase portraits in Figure 6.16 and the transplantation of a chessboard in Figure 6.17. We encourage the reader to work out the details.

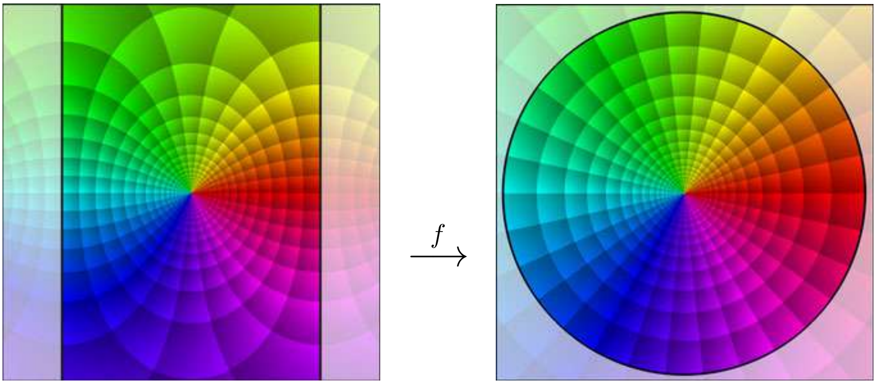


Figure 6.16: Phase portrait of the tangent function in the strip $-\pi/4 < \operatorname{Re} z < \pi/4$

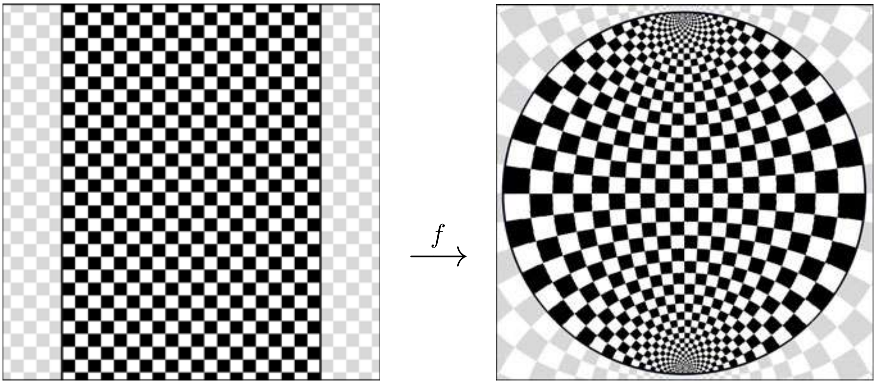


Figure 6.17: Mapping an infinite strip to the disk by the tangent function

Example 6.2.6 (The Joukovski Mapping). The left window of Figure 6.18 depicts the phase portrait of the rational function which is defined on $\widehat{\mathbb{C}}$ by

$$f(z) = \frac{1}{2}\left(z + \frac{1}{z}\right) \tag{6.12}$$

with $f(\infty) := \infty$. It does not immediately disclose its relation to conformal mappings, but indeed the restriction of f to the exterior $\mathbb{E} := \{z \in \widehat{\mathbb{C}} : |z| > 1\}$ of the unit circle maps $\mathbb{E}$ conformally onto the *slit domain* $S := \widehat{\mathbb{C}} \setminus [-1, 1]$ as illustrated in Figure 6.19.

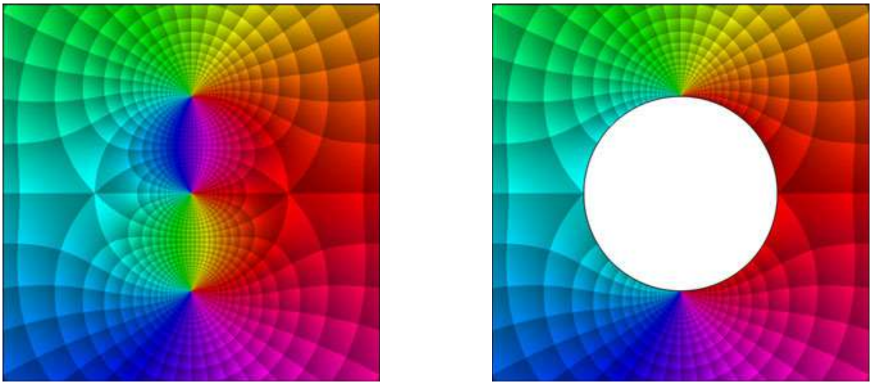


Figure 6.18: Phase portraits of the function f and of the Joukovski mapping

This function is called the *Joukovski mapping*, its phase portrait is shown in Figure 6.18 (right). Note that $\mathbb{E}$ (including the point at infinity) and S are simply connected subdomains of $\widehat{\mathbb{C}}$.

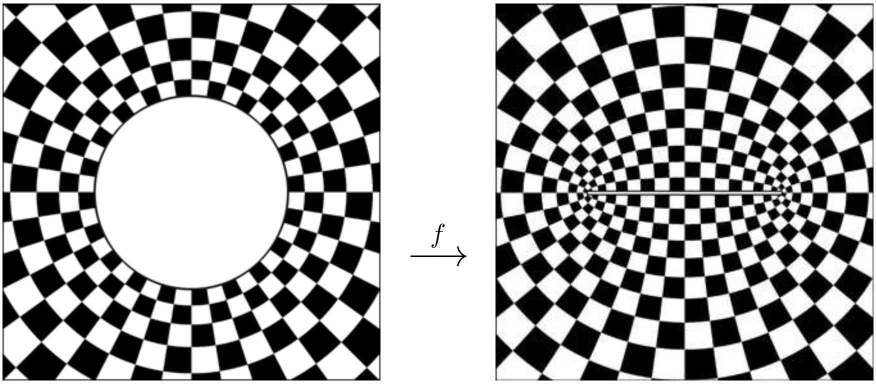


Figure 6.19: The Joukovski mapping of $\mathbb{E}$ onto a slit domain

The mapping properties of f can be understood more clearly from the phase portrait of the inverse function $f^{-1}(w) = w + \sqrt{w^2 - 1}$, depicted in Figure 6.20 (right).

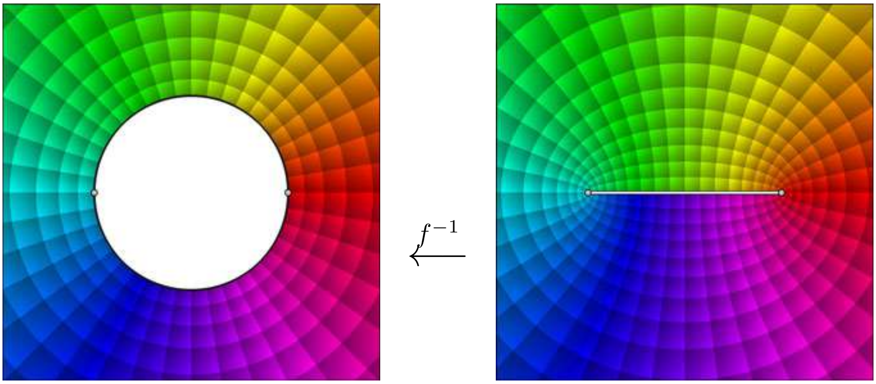


Figure 6.20: Phase portrait of the inverse Joukovski mapping (right picture)

The images of the circular contour lines in the z-plane on the left-hand side look like ellipses in the w-plane. It can be easily verified that these are indeed confocal ellipses with foci -1 and 1, and that the radial phase contour lines are mapped onto an orthogonal family of hyperbolas. Since the function f defined by (6.12) is invariant with respect to the substitution $z \mapsto 1/z$, it also maps the unit disk $\mathbb{D}$ conformally onto S. This is visualized in Figure 6.21, where the two domains $\mathbb{D}$ and $\mathbb{E}$ are depicted in the middle, the image of $\mathbb{E}$ is shown on the left, and the image of $\mathbb{D}$ is on the right.

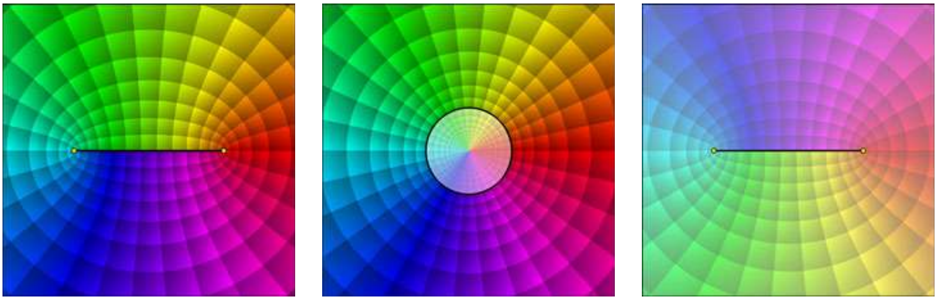


Figure 6.21: The Joukovsky mapping in the exterior and the interior of $\mathbb{T}$

An interesting application of the Joukovski map to flow problems is illustrated in Figure 6.22. The picture on the right shows the stream lines of a (potential) flow in the w-plane slit along the segment $[-1, 1]$. Since the fluid moves in the direction parallel to the slit (say from left to right), the flow is not disturbed by the obstacle.

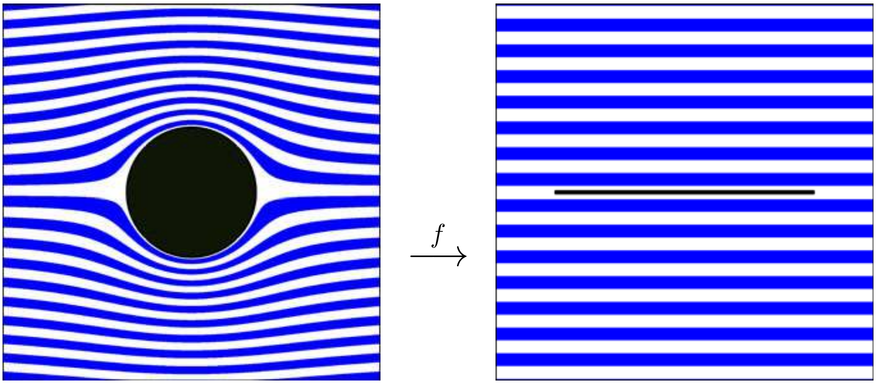


Figure 6.22: The Joukovsky mapping and potential flow around a circular cylinder

The picture on the left is the result of pulling back the stream lines from the w-plane to the z-plane by the Joukovski mapping $w = f(z)$. It resembles a plane flow around the black disk – the background of this observation has been explained in Section 4.6.

A remarkable variety of conformal mappings can already be constructed by combinations of the functions (and their inverses) from the above examples.

Example 6.2.7. The function $w = f(z) := \mathrm{i}\sqrt{1-z^2}$ maps the upper half-plane to the upper half-plane slit along the segment $[0, \mathrm{i}]$ as illustrated in Figure 6.23.

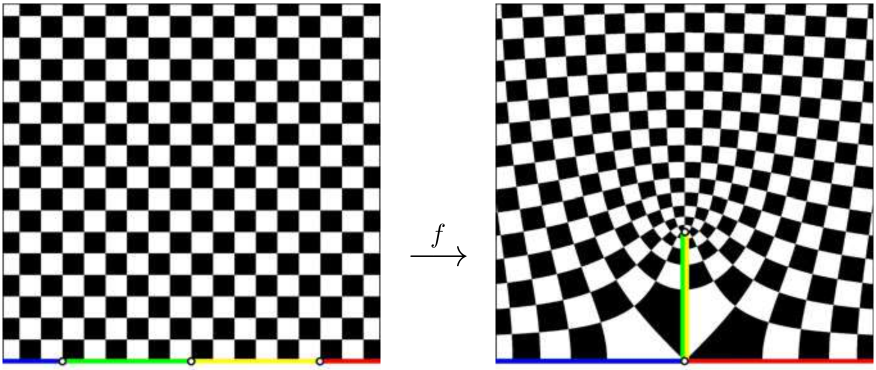


Figure 6.23: Conformal mapping of a half-plane to a slit half-plane

Basically, the mapping f is a composition of four functions: first $f_1 : z \mapsto -z^2$ maps the upper half-plane to the full plane slit along the negative real axis. Then $f_2 : z \mapsto z + 1$ shifts the plane one unit to the right, then the principal branch[1] of the square root $f_3 : z \mapsto \sqrt{z}$ ‘folds’ the plane onto the right half-plane by opening up part of the slit (the red and blue lines) along the negative real axis. Finally, $f_4 : z \mapsto \mathrm{i}\, z$ rotates everything by an angle of $\pi/2$.

---

[1] This is the reason for *not* defining f by $\sqrt{z^2-1}$, otherwise we could not have used the principle branch.

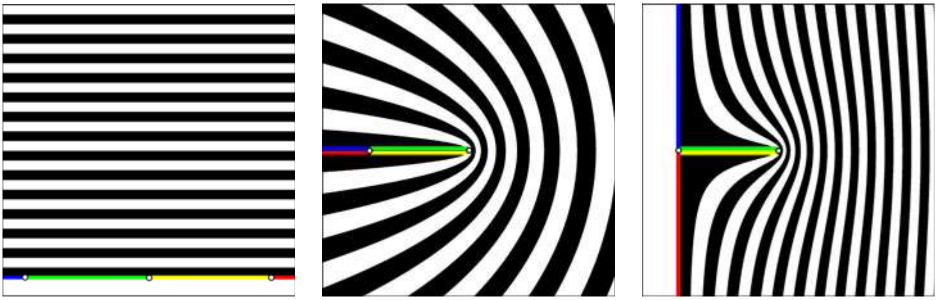


Figure 6.24: Successive construction of the function $f(z) = \mathrm{i}\sqrt{1-z^2}$

Figure 6.24 shows a strip-pattern in the original plane (left), and its images after the action of f_1 (middle), and $f_3 \circ f_2 \circ f_1$ (right), respectively.

Example 6.2.8. The domain $G := \{z \in \mathbb{C} : \operatorname{Im} z > 0,\ |z - \mathrm{i}| > 1\}$ is the upper half-plane minus a disk tangent to the real line. It is mapped onto the upper half-plane by either of the functions f and g, defined by

$$f(z) := \mathrm{e}^{-2\pi/z}, \qquad g(z) := \pi \coth \frac{\pi}{z} = \pi\, \frac{\mathrm{e}^{\pi/z} + \mathrm{e}^{-\pi/z}}{\mathrm{e}^{\pi/z} - \mathrm{e}^{-\pi/z}}.$$

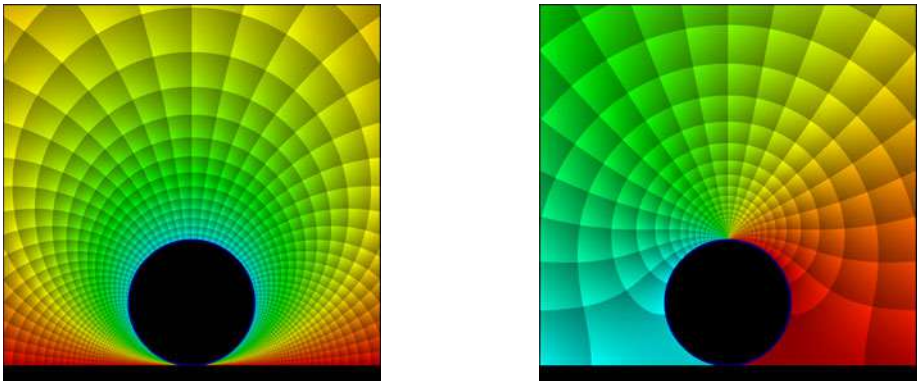


Figure 6.25: Enhanced phase portraits of the functions f (left) and g (right)

The phase portraits of f and g are shown in Figure 6.25. Figure 6.26 depicts the result of pulling back ‘stream lines’ of a flow parallel to the x-axis (shown on the right-hand side of Figure 6.22) from the upper half-plane to G using the functions f and g, respectively. While the result in the picture on the right looks plausible, the exotic stream lines in the picture on the left certainly do not satisfy our expectations from a realistic flow. These differences are caused by the normalization of the conformal mappings. While $f(z) \to 1$ as $z \to \infty$, the function g has the asymptotic behavior $g(z) \sim z$ at infinity. Consequently the parallel flow at some distance of the origin is reproduced almost undisturbed by g, but not by f.

Note that the wild behavior on the left side of the black disk has nothing to do with turbulence, it is just an optical effect, caused by the very dense stream lines and the limited resolution of the graphics.

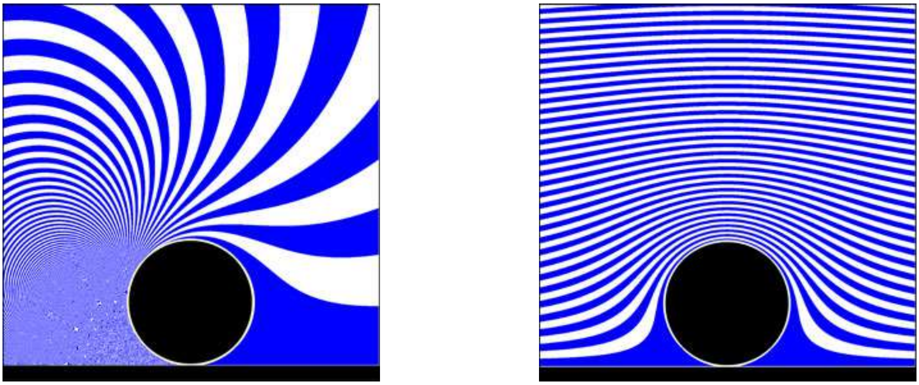


Figure 6.26: Stream lines of potential flows generated by f (left) and g (right)

6.3 Möbius Transformations

In this section we study conformal mappings on the Riemann sphere. As the point at infinity will often be involved, the notion of analytic functions is usually understood in the wider sense of Definition 3.5.1.

As we have seen in the preceding section, the conformal automorphisms of the Riemann sphere are precisely the *Möbius transformations*

$$f(z) = \frac{az+b}{cz+d}, \qquad a,b,c,d \in \mathbb{C},\ ad-bc \neq 0. \tag{6.13}$$

Recall from Section 2.3 that any Möbius transformation is a composition of translations $z \mapsto z+b$, dilations $z \mapsto az$ $(a \in \mathbb{R}_+)$, rotations $z \mapsto az$ $(a \in \mathbb{T})$, and the (anti-)inversion $z \mapsto 1/z$. This implies that Möbius transformations map circles to circles. Speaking of circles, we always mean circles on the Riemann sphere, unless the contrary is explicitly mentioned.

Figure 6.27 visualizes the Möbius transformation $f(z) = (z+\mathrm{i})/(z-1)$ on the Riemann sphere. The left picture is an enhanced phase portrait, the chessboard coloring on the right-hand side emphasizes the conformality of the mapping.

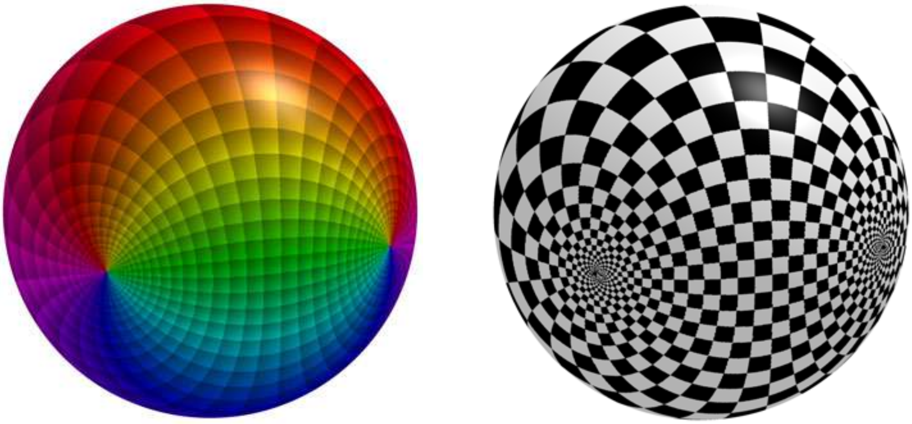


Figure 6.27: Visualization of a Möbius transformation on the sphere

Matrix Representations. Möbius transformations form a group with respect to composition. Calculations can be reduced to matrix algebra by associating the Möbius transformations involved with 2×2 matrices. If

$$f(z) = \frac{az+b}{cz+d} \asymp M_f := \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \qquad g(z) = \frac{Az+B}{Cz+D} \asymp M_f := \begin{bmatrix} A & B \\ C & D \end{bmatrix},$$

then the entries of the matrix product $M_g\, M_f$ furnish coefficients of the Möbius transformation $g \circ f$. Analogously, the inverse Möbius transformation f^{-1} can be constructed from the inverse matrix M_f^{-1}. After cancelling the determinant of M_f we get the representation

$$f^{-1}(z) = \frac{dz-b}{-cz+a}. \tag{6.14}$$

Although any 2×2 matrix M_f with entries a, b, c, d and $\det M_f = ad - bc \neq 0$ determines a Möbius transformation, this correspondence is not injective because we get the same associated Möbius transformation if all entries of the matrix are multiplied by a common non-zero factor. We therefore use a *normalized* representation, where the entries are scaled such that $ad - bc = 1$. The entries of a normalized matrix representation are unique up to a factor of -1.

Normalized matrix representations for the translation $z \mapsto z + b$, the roto-stretch $z \mapsto az$, and the inversion $z \mapsto 1/z$ are (in this order)

$$\begin{bmatrix} 1 & b \\ 0 & 1 \end{bmatrix}, \qquad \begin{bmatrix} a^{1/2} & 0 \\ 0 & 1/a^{1/2} \end{bmatrix}, \qquad \begin{bmatrix} 0 & \mathrm{i} \\ \mathrm{i} & 0 \end{bmatrix}. \tag{6.15}$$

Fixed Points. In the context of conformal mappings one often encounters another representation of Möbius transformations. It uses the images $f(z_1), f(z_2), f(z_3)$ of three distinct points z_1, z_2, z_3 on the sphere, for example $z_1 = 1$, $z_2 = 0$ and $z_3 = \infty$. Before describing this representation in detail, we first derive an auxiliary result. Recall that $z \in \widehat{\mathbb{C}}$ with $f(z) = z$ is said to be a *fixed point* of f.

Lemma 6.3.1. *Any Möbius transformation different from the identity has at most two fixed* fixed points. *If $f(z) = (az + b)/(cz + d)$ with $ad - bc = 1$, then f has exactly one fixed point if $a + d = \pm 2$, and exactly two fixed points otherwise.*

Proof. In the following we assume that f is normalized by $ad - bc = 1$.

1. If $c = 0$, then $d \neq 0$ and $f(z) = (az + b)/d$ is a linear function. The fixed point equation $f(z) = z$ has no solution $z \in \mathbb{C}$ if $d = a$, and exactly one such solution $z = b/(d - a)$, otherwise. Because $ad - bc = 1$, the first case appears if and only if $d = a = 1$ or $d = a = -1$. In both cases f also has a fixed point at infinity.

2. If $c \neq 0$, then neither $z = \infty$ nor $z = -d/c$ are fixed points of f. For all other points z the fixed point condition $f(z) = z$ is equivalent to

$$c\,z^2 - (a - d)\,z - b = 0. \tag{6.16}$$

If $a + d \neq \pm 2$, this quadratic equation has two single solutions,

$$z_{1/2} = \frac{1}{2c}\left((a - d) \pm \sqrt{(a + d)^2 - 4}\right),$$

while it has one (double) solution $z_1 = (a - d)/(2c)$ if $a + d = \pm 2$. □

Inspection of the proof yields the following corollary which we state here for future reference.

Corollary 6.3.2. *If a Möbius transformation f has only the fixed point ∞, then $f(z) = z + b$ with $b \in \mathbb{C} \setminus \{0\}$. If f has fixed points 0 and ∞, then $f(z) = az$ with $a \in \mathbb{C} \setminus \{0, 1\}$.*

The Cross-Ratio. We now introduce a notion which is not only useful in the context of Möbius transformations, but has many applications in plane geometry.

Definition 6.3.3. Let z_0, z_1, z_2, z_3 be points on the sphere $\widehat{\mathbb{C}}$ and assume that the set $\{z_0, z_1, z_2, z_3\}$ contains at least three distinct points. Then

$$[z_0, z_1; z_2, z_3] := \frac{z_0 - z_2}{z_0 - z_3} : \frac{z_1 - z_2}{z_1 - z_3} \tag{6.17}$$

is called the *cross-ratio* of z_0, z_1, z_2, z_3.

With the usual conventions about the arithmetic operations involving zero and infinity, the double fraction on the right-hand side can always be rewritten so that it is well defined. Note that there is no general agreement about the definition of the cross-ratio in the literature, but all variants can be obtained from (6.17) by appropriate permutation of the entries z_0, z_1, z_2, z_3.

The meaning of definition Definition 6.3.3 becomes clear when we assume that z_1, z_2, z_3 are distinct and consider z_0 as a variable point: the function

$$f(z) := [z, z_1; z_2, z_3] \tag{6.18}$$

is a Möbius transformation which maps z_1 to 1, z_2 to 0, and z_3 to ∞. In fact the cross-ratio (6.18) is the only Möbius transformation with this property: if g is any Möbius transformation with $g(z_k) = f(z_k)$ for $k = 1, 2, 3$, then $h := f^{-1} \circ g$ is a Möbius transformation with three different fixed points z_1, z_2, z_3. By Lemma 6.3.1 h must be the identity.

Theorem 6.3.4. *Let z_1, z_2, z_3 and w_1, w_2, w_3 be two triples of mutually distinct points on the Riemann sphere. Then there exists a unique Möbius transformation f with $f(z_k) = w_k$ for $k = 1, 2, 3$.*

Proof. The Möbius transformation $g(z) := [z, z_1; z_2, z_3]$ maps $z_1 \mapsto 1$, $z_2 \mapsto 0$, $z_3 \mapsto \infty$, and $h(w) := [w, w_1; w_2, w_3]$ maps $w_1 \mapsto 1$, $w_2 \mapsto 0$, $w_3 \mapsto \infty$. Then $f := h^{-1} \circ g$ solves the problem. Uniqueness has already been shown. □

Corollary 6.3.5. *The cross-ratio is an invariant of any Möbius transformation f,*

$$[f(z_0), f(z_1); f(z_2), f(z_3)] = [z_0, z_1; z_2, z_3]. \tag{6.19}$$

Proof. We may assume that z_1, z_2, z_3 are distinct. Let f be a Möbius transformation and set $w_k := f(z_k)$ for $1, 2, 3$. Then f has the same mapping properties as in Theorem 6.3.4, and with the notation used in its proof we have $(h \circ f)(z) = g(z)$ for all z. Setting $z := z_0$ we obtain (6.19). □

The cross-ratio has a number of interesting applications in plane geometry (see [22], for instance). Here is one of them.

Theorem 6.3.6. *The cross-ratio of four points on the Riemann sphere is real if and only if all points lie on a circle.*

Proof. 1. Denote the four points by z_0, z_1, z_2, z_3. If $z_1 = -1, z_2 = 0, z_3 = 1$ it is geometrically plausible that $[z_0, z_1; z_2, z_3] = 2z_0/(z_0 + 1)$ is real if and only if z_0 lies on the extended real line.

2. In the general case, there exists a Möbius transformation f which maps z_1, z_2, z_3 to $-1, 0, 1$, respectively (see Theorem 6.3.4). Since f and f^{-1} map circles to circles, the points z_0, z_1, z_2, z_3 lie on a circle if and only if this holds for their image points $f(z_0), -1, 0, 1$. Now the assertion follows from the invariance property $[z_0, z_1; z_2, z_3] = [f(z_0), -1; 0, 1]$ and the first step. □

The circle $C(z_1, z_2, z_3)$ through the points z_1, z_2, z_3 divides the Riemann sphere into two components (spherical caps), $D_+(z_1, z_2, z_3)$ and $D_-(z_1, z_2, z_3)$.

From Theorem 6.3.6 the sign of the imaginary part of the cross-ratio $[z, z_1; z_2, z_3]$ is constant in every component and we choose the plus-minus-labelling so that

$$D_{\pm}(z_1, z_2, z_3) := \{z \in \widehat{\mathbb{C}} : \pm \mathrm{Im}\,[z, z_1; z_2, z_3] > 0\}.$$

The domains $D_-(z_1, z_2, z_3)$ and $D_+(z_1, z_2, z_3)$ are called the *left side* and the *right side* of $C(z_1, z_2, z_3)$, respectively. This definition is geometrically motivated: when we walk along the boundary circle from z_1 via z_2 to z_3, then D_- lies on our left, while D_+ lies on the right. Here we consider $C(z_1, z_2, z_3)$ as an *oriented circle*.

Theorem 6.3.7. *If a Möbius transformation f maps z_1, z_2, z_3 to w_1, w_2, w_3, respectively, then it maps the domain $D_-(z_1, z_2, z_3)$ onto $D_-(w_1, w_2, w_3)$, and the domain $D_+(z_1, z_2, z_3)$ onto $D_+(w_1, w_2, w_3)$.*

The proof is left as an exercise. Figure 6.28 visualizes a Möbius transformation which maps the interior of a circle conformally onto the exterior of the image circle. Both domains are left sides of the corresponding circles $C(z_1, z_2, z_3)$ and $C(w_1, w_2, w_3)$, respectively.

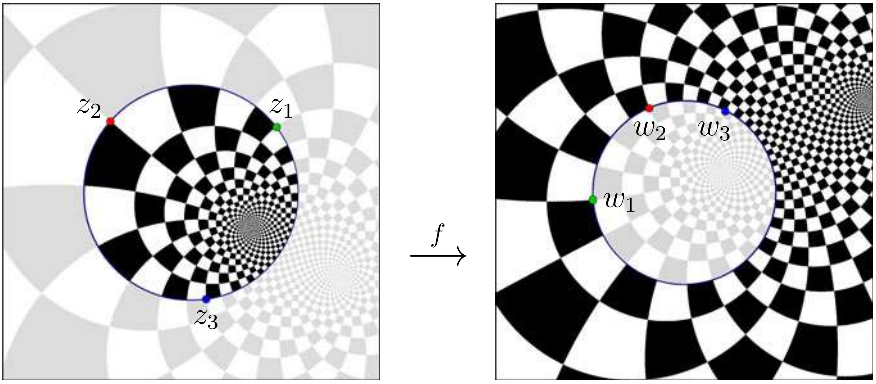


Figure 6.28: Orientation and mapping properties of a Möbius transformation

Dynamical Systems. Another way to deepen our understanding of Möbius transformations is to consider them in the framework of *discrete dynamical systems*. Here "discrete dynamical system" is just another name for a mapping f of the Riemann sphere onto itself, but we prefer to use this terminology since we are interested in some specific questions which are typically posed in this setting. In particular we shall explore "how points move" when f is applied once, twice, and repeatedly over and over again.

In order to investigate this problem, we define the *iterates* of a function $f : \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ by $f^0(z) := z$ (f_0 is the identity function), $f^1 := f$ and $f^{k+1} = f \circ f^k$ for $k = 1, 2, \ldots$. If f is invertible, we further set $f^{-k} := (f^{-1})^k$. Then the *orbit* of a point z_0 on $\widehat{\mathbb{C}}$ is the set

$$O(z_0) := \{f^k(z_0) : \quad k \in \mathbb{Z}\}.$$

Though orbits are discrete sets, they are often arranged along certain curves. If, for example, f is a translation, say $f(z) = z + 1$, then every orbit $O(z_0)$ lies on the 'horizontal' line $\operatorname{Im} z = \operatorname{Im} z_0$ through z_0. If z_0 is on such a line, then $f^k(z_0)$ stays on it, for all $k \in \mathbb{Z}$.

Invariant Sets. This last observation motivates the concept of invariant sets: a set S is called *invariant* with respect to the action of f (or simply *invariant for* f) if $f(S) \subset S$. If S is a Jordan curve or a Jordan arc we also say that f *acts along* S. The simplest invariant sets of a mapping are its *fixed points*. According to Lemma 6.3.1, any Möbius transformation, which is not the identity, has one or two of them.

Let us first assume that f has only *one fixed point* z_1. If $z_1 = \infty$, then, by Corollary 6.3.2, $f(z) = z + b$ with $b \in \mathbb{C} \setminus \{0\}$. If $z_1 \neq \infty$, then the Möbius transformation

$$g(z) := 1/(z - z_1) \tag{6.20}$$

sends z_1 to ∞, so that $F := g \circ f \circ g^{-1}$ has exactly one fixed point, which is at infinity. Consequently, we must have $F(z) = z + b$.

Defining g by (6.20) if $z_1 \neq \infty$ and by $g(z) := z$ if $z_1 = \infty$, we obtain a representation which is valid in both cases,

$$f = g^{-1} \circ F \circ g, \qquad F(z) = z + b. \tag{6.21}$$

If f has *two fixed points* z_1, z_2, we can choose the indexing so that $z_1 \neq \infty$. If $z_2 = \infty$ we set $g(z) := z - z_1$, otherwise we define g by $g(z) := (z - z_1)/(z - z_2)$. In both cases the Möbius transformation g sends z_1 to 0 and z_2 to ∞, so that $F := g \circ f \circ g^{-1}$ is a Möbius transformation with fixed points 0 and ∞. By Corollary 6.3.2, we must have $F(z) = \kappa\, z$ with $\kappa \in \mathbb{C} \setminus \{0, 1\}$, so that we get the representation

$$f = g^{-1} \circ F \circ g, \qquad F(z) = \kappa\, z. \tag{6.22}$$

The constant κ is called a *multiplier* for the Möbius transformation f. A short calculation yields that

$$\kappa = f'(z_1) = 1/f'(z_2) \qquad (z_2 \neq \infty).$$

If the fixed points z_1 and z_2 are interchanged, κ is replaced by $1/\kappa$, which is the reason why not only κ but also $1/\kappa$ is said to be a multiplier for f.

The Möbius transformations f and F in (6.21) and (6.22) are related to each other by $f = g^{-1} \circ F \circ g$ for some Möbius transformation g. If such a representation exists, we say that f and F are *conjugate*. It is convenient to think of f as a mapping in the z-plane, while F acts in another Z-plane. Then g is a bijective mapping from the z-plane to the Z-plane which transplants the action of f from the z-plane to the Z-plane.

Many properties of Möbius transformations are preserved under conjugation. In particular, a point z_0 is a fixed point of f if and only $g(z_0)$ is a fixed point of F, and a set S is invariant for f if and only if $g(S)$ is invariant for F.

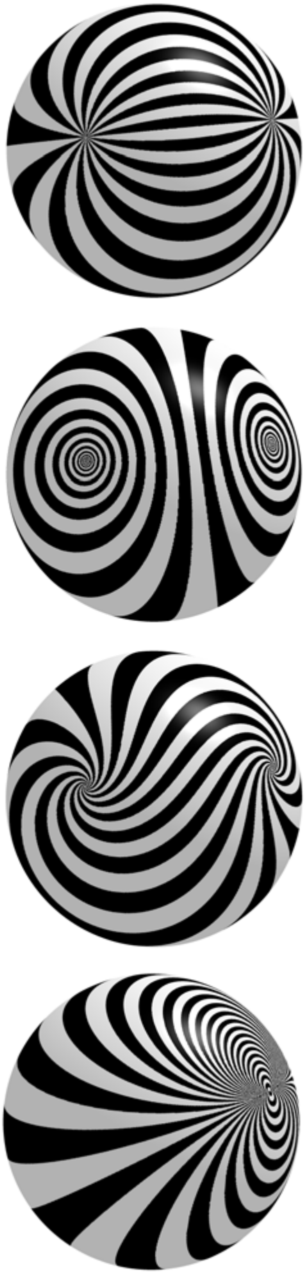


Figure 6.29: The four classes of Möbius transformations

In the following we apply this idea in the reverse direction: starting with invariant sets S for the simple mapping F we obtain invariant sets for the conjugate Möbius transformation f as the preimages $g^{-1}(S)$.

1. If f has two fixed points and the multiplier κ in (6.22) is *real and positive*, then $F(Z) = \kappa Z$ is a *dilation* which acts along radial rays through the origin in the Z-plane. The mapping $z = g^{-1}(Z)$ sends these rays to circular arcs in the z-plane which connect the two fixed points $z_1 = g^{-1}(0)$ and $z_2 = g^{-1}(\infty)$. Möbius transformations of this kind are called *hyperbolic*.

2. If the multiplier $\kappa \neq 1$ in (6.22) is *unimodular*, $\kappa \in \mathbb{T} \setminus \{1\}$, then F is a *rotation*, which acts along circles centered at the origin in the Z-plane. In the z-plane, these circles are *orthogonal* to the circles through the two fixed points z_1 and z_2. The corresponding Möbius transformations are called *elliptic*.

3. In all other cases with two fixed points the mapping F is a *rotostretch*, which acts along *logarithmic spirals* centered at the origin. Since these spirals intersect all radial rays in the Z-plane at the same angle, their images in the z-plane have the same property with respect to the circles through z_1 and z_2. Such curves are called “loxodromes”, and we therefore speak of *loxodromic* Möbius transformations.

4. If f has only one fixed point, the mapping $F(Z) = Z + b$ in the representation (6.21) acts along parallel lines in the direction of the vector b in the Z-plane. These invariant sets for F are mapped by $Z \mapsto z = g^{-1}(Z)$ onto a family of mutually tangent circles through z_1 which are invariant for f. Möbius transformations of this type are said to be *parabolic*.

Classification. We summarize the result in the following theorem which classifies normalized Möbius transformations explicitly in terms of their coefficients. For computational details we refer to Palka [52], Section IX.2.

Theorem 6.3.8. *The normalized Möbius transformation* $f(z) = (az+b)/(cz+d)$ *with* $ad - bc = 1$ *is* parabolic *if and only if* $a + d = \pm 2$*, it is* hyperbolic *if and only if* $a+d$ *is real with* $|a+d| > 2$*, it is* elliptic *if and only if* $a+d$ *is real with* $|a+d| < 2$*, and it is* loxodromic *if and only if* $a+d$ *is not real.*

In the non-parabolic case, the representation (6.22) can be rewritten as

$$\frac{f(z)-z_1}{f(z)-z_2} = \kappa\,\frac{z-z_1}{z-z_2}, \qquad \text{or} \qquad f(z)-z_1 = \kappa\,(z-z_1), \tag{6.23}$$

depending on whether $z_2 \neq \infty$ or $z_2 = \infty$, respectively. These formulas are often referred to as the *multiplier fixed point form* of f.

From (6.23) or its abstract version we can immediately derive a similar representation for the iterates f^k,

$$\frac{f^k(z)-z_1}{f^k(z)-z_2} = \kappa^k\,\frac{z-z_1}{z-z_2}, \qquad \text{or} \qquad f^k(z)-z_1 = \kappa^k\,(z-z_1). \tag{6.24}$$

Solving these equations with respect to $f^k(z)$ we can obtain explicit formulas for the iterates f^k, but this is not necessary in order to understand the "long-time behavior" of the dynamical system, if $|\kappa| < 1$ and $z \neq z_2$, the right-hand sides of (6.24) tend to zero as $k \to +\infty$, so $f^k(z_0)$ tends to z_1. Similarly, if $z \neq z_1$, then $f^k(z)$ approaches z_2 as $k \to -\infty$. The fixed point z_1 is *attracting* while z_2 is *repelling*. If $|\kappa| > 1$, the roles of z_1 and z_2 are reversed. The reader is encouraged to consider the parabolic case and the case when $|\kappa| = 1$.

6.4 The Riemann Mapping Theorem

In this section we consider the general existence problem for conformal mappings between simply connected domains. The following celebrated result is not only a miracle in complex function theory but is also incredibly beautiful and unbeatable in its generality. It is not without reason that many consider it to be one of the "eternal" theorems in mathematics.

Theorem 6.4.1 (Riemann Mapping Theorem)**.** *Any simply connected domain* G*, which is a proper subset of the complex plane, is conformally equivalent to the unit disk* $\mathbb{D}$*. More precisely, for any* $z_0 \in G$ *there exists a unique analytic function* g *which maps* G *conformally onto* $\mathbb{D}$ *and satisfies the normalization conditions*

$$g(z_0) = 0, \qquad g'(z_0) > 0. \tag{6.25}$$

Given the depth of this result, its proof is surprisingly simple, at least when the necessary machinery has already been established. Our main tool will be Montel's theorem (Theorem 5.2.4).

Proof. 1. The following approach, due to Constantin Carathéodory, is based on an *extremal principle* and a *normal families argument.* To begin with, we shall verify that the set

$$\mathcal{F} := \{f : G \to \mathbb{D} : \; f \text{ analytic, injective, } f(z_0) = 0\} \tag{6.26}$$

is a non-void normal family. Then we pick a point $p \in G \setminus \{z_0\}$ and show that $\mathcal{F}$ contains functions $g \in \mathcal{F}$ which satisfy the *extremal condition*

$$|g(p)| = \max\{|f(p)| : \; f \in \mathcal{F}\}. \tag{6.27}$$

Finally, we prove that any such *extremal function* g maps G conformally onto D.

2. In order to check that $\mathcal{F}$ is not empty, we distinguish three cases of increasing difficulty.

2.1. If G is *bounded*, there exists a linear function $f(z) = az+b$ with $f(z_0) = 0$ that maps G onto a subset of $\mathbb{D}$ and thus belongs to $\mathcal{F}$.

2.2. If $\mathbb{C} \setminus G$ *contains an inner point* a, the mapping $z \mapsto 1/(z-a)$ sends G to a simply connected bounded domain, which brings us back to 2.1.

2.3. Finally, in the *general case*, let $a \in \mathbb{C} \setminus G$. Then the function $z \mapsto z - a$ does not vanish in G, and since G is simply connected, there exists an analytic branch of the logarithm $z \mapsto \log(z-a)$ on G (see Lemma 3.6.11). The function h defined by $h(z) := \exp\big(\frac{1}{2}\log(z-a)\big)$ is an analytic branch of the square root of $z - a$, that is $h(z)^2 = (z-a)$ for all $z \in G$. It is clear that h is injective, so that it maps G conformally onto a simply connected domain $h(G)$. Moreover $h(z) = -h(w)$ with $z, w \in G$ implies that

$$z - a = \big(h(z)\big)^2 = \big(-h(w)\big)^2 = w - a,$$

from which we successively conclude that $z = w$, $h(z) = -h(z)$, $h(z) = 0$, and $a = z \in G$, which is impossible. Consequently the sets $h(G)$ and $-h(G)$ are disjoint, and because $-h(G)$ is a non-empty simply connected domain, $\mathbb{C} \setminus h(G)$ contains inner points. So we again return to the situation considered in Step 2.2.

3. Since the family $\mathcal{F}$ is not empty, $s := \sup\{|f(p)| : \; f \in \mathcal{F}\} > 0$, and we can choose a sequence $(f_n) \subset \mathcal{F}$ with $|f_n(p)| \to s$. The functions f_n are uniformly bounded, hence, by Montel's theorem (Theorem 5.2.4), a subsequence of f_n converges normally to a function g which is analytic in G. This limit function satisfies $g(z_0) = 0$, $|g(p)| = s$, and $|g(z)| \le 1$ for all $z \in G$. Because $s \ne 0$, g is not constant, so that the maximum principle (Theorem 3.4.3) implies $|g(z)| < 1$ for all $z \in G$. Moreover, it follows from Hurwitz' theorem (or more precisely, from Corollary 5.1.6) that g is injective. Thus we have shown that g is an extremal function in $\mathcal{F}$.

4. In order to verify that g is a conformal mapping of G onto $\mathbb{D}$ we prove that any extremal function g in $\mathcal{F}$ is a surjective mapping of G onto $\mathbb{D}$. The following construction shows that g cannot be extremal if $D := g(G)$ is a proper subset of

$\mathbb{D}$, since then there exists a function $f \in \mathcal{F}$ with $|f(p)| > |g(p)|$. For convenience, the mapping properties of the functions involved in the proof are visualized in Figure 6.30.

So, let us assume that there exists a point $a \in \mathbb{D} \setminus D$. Then the Möbius transformation

$$\varphi_a(z) = \frac{z-a}{1-\bar{a}z} \qquad (z \in D),$$

maps D conformally onto a simply connected domain $D_a \subset \mathbb{D}$. Since $\varphi_a(a) = 0$, we conclude that $0 \notin D_a$, and as in Step 2.3 we can choose an analytic branch σ of the square root which maps D_a conformally onto a simply connected domain $D_b \subset \mathbb{D}$. Note that there are exactly two such branches, σ and $-\sigma$. Now we set $b := \sigma(a)$ and consider the Möbius transformation

$$\varphi_b(z) = \frac{z-b}{1-\bar{b}z} \qquad (z \in D_b),$$

which maps D_b conformally onto a simply connected domain D_c. The composition $\psi_c := \varphi_b \circ \sigma \circ \varphi_a$ maps D conformally onto D_c and satisfies $\psi_c(0) = 0$, so that $f := \psi_c \circ g \in \mathcal{F}$.

5. In the final step we prove that $|f(p)| > |g(p)|$, which shows that g is not extremal. The crucial observation is that $\psi_c^{-1} : D_c \to D$ extends to an analytic function $\Psi_c : \mathbb{D} \to \mathbb{D}$ with $\Psi_c(0) = 0$. Indeed we have $\psi_c^{-1} = \varphi_a^{-1} \circ \sigma^{-1} \circ \varphi_b^{-1}$, where φ_a^{-1}, σ^{-1}, and φ_b^{-1} are the restrictions of

$$\Phi_a(z) := \frac{z+a}{1+\bar{a}z}, \qquad S(z) := z^2, \qquad \Phi_b(z) := \frac{z+b}{1+\bar{b}z} \qquad (z \in \mathbb{D}),$$

to D_a, D_b, and D_c, respectively. Consequently, ψ_c^{-1} is the restriction to D_c of a Blaschke product of order two, namely

$$\Psi_c(z) = \big(\Phi_a \circ S \circ \Phi_b\big)(z) = z\,\frac{z-c}{1-\bar{c}z}, \qquad z \in \mathbb{D},$$

where $c := \varphi_b(-b) = -2b/(1+|b|^2) \in \mathbb{D}$. Because Ψ_c maps $\mathbb{D}$ into $\mathbb{D}$, satisfies $\Psi_c(0) = 0$, and since Ψ_c is not a rotation, Schwarz' lemma (Theorem 3.4.12) tells us that $|\Psi_c(z)| < |z|$ for all $z \in \mathbb{D} \setminus \{0\}$. Inserting $z := f(p)$ we get that $|g(p)| < |f(p)|$.

6. In order to show uniqueness of the normalized mapping function, we assume that f and g are two conformal mappings of G onto $\mathbb{D}$ which satisfy the extra conditions (6.25). Then $h := g \circ f^{-1}$ is a conformal automorphism of $\mathbb{D}$ with $h(0) = h^{-1}(0) = 0$ and $h'(0) > 0$. By Theorem 6.2.3, h is the identity mapping of $\mathbb{D}$, that is $f = g$. □

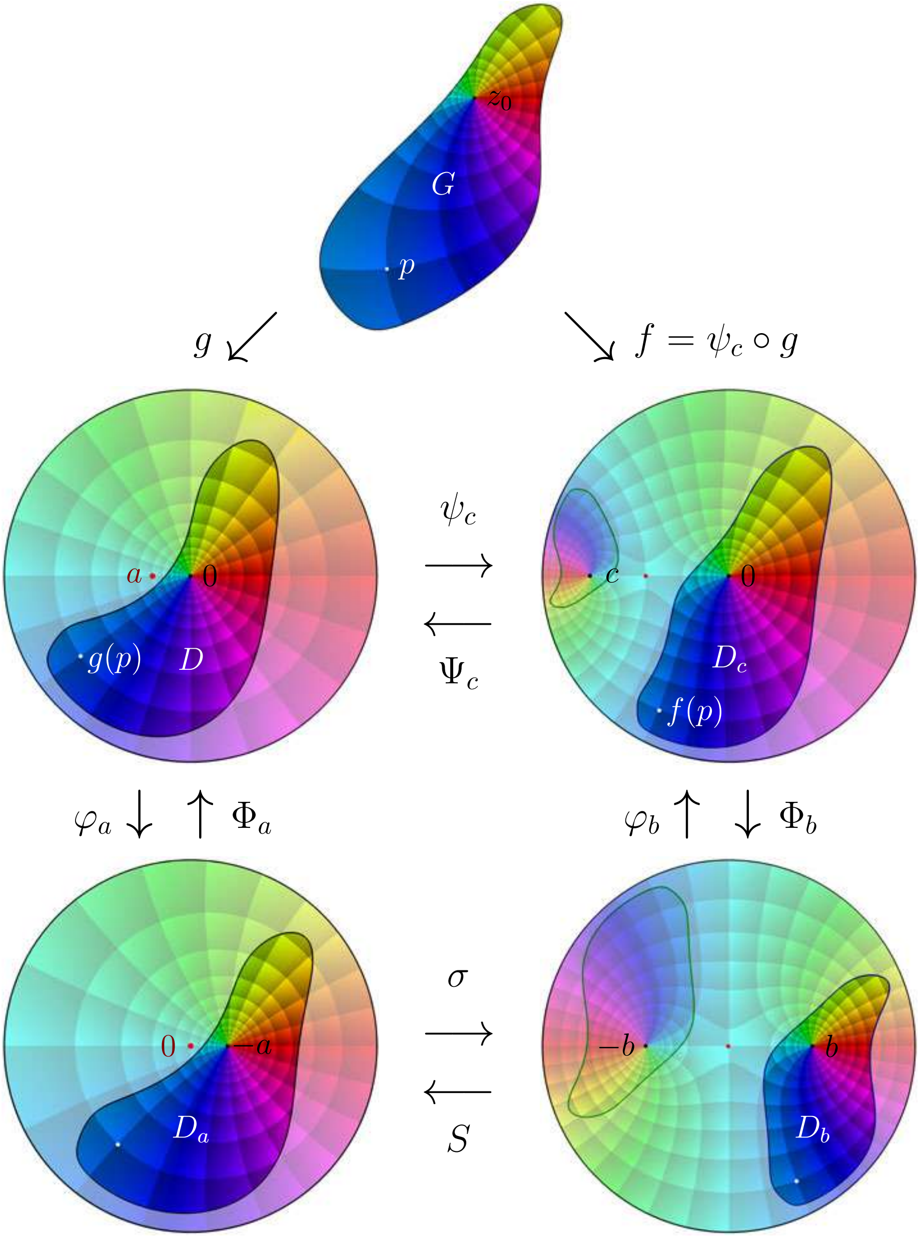


Figure 6.30: Illustration of the proof of the Riemann mapping theorem

Figure 6.30 illustrates the main idea of the proof. The colorings of the domains are transplantations of the coloring of the unit disk $\mathbb{D}$ of the picture on the left in the second row. The picture on top is the phase portrait of g. Functions denoted by small letters map between the domains with saturated colors, functions denoted by capitals map $\mathbb{D}$ onto $\mathbb{D}$. In the two pictures in the right column, the image domains with respect to the second branch $-\sigma$ of the square root are also shown.

A More General Extremal Principle. We remark that the conformal mappings with $f(z_0) = 0$ remain the extremal functions for $|f(p)| \to \text{Max}$ even when $\mathcal{F}$ is replaced by the larger set

$$\mathcal{G} := \{f : G \to \mathbb{D} : \ f \text{ analytic}, \ \ f(z_0) = 0\}, \tag{6.28}$$

where the *a-priori* assumption of injectivity has been abandoned. For a proof it suffices to consider the compositions $f \circ g : \mathbb{D} \to \mathbb{D}$, where $f \in \mathcal{G}$ and g is a conformal mapping of $\mathbb{D}$ onto G, and to apply Schwarz' lemma. This version of Carathéodory's extremal principle is the nucleus of a more general result which characterizes solutions to nonlinear boundary value problems (so-called Riemann-Hilbert problems) by extremal properties. Interested readers are referred to Chapter 3 of Wegert [68].

Topological Versus Conformal Equivalence. The remarkable generality of the Riemann mapping theorem manifests itself when it is rephrased in the language of topology. Recall that two sets X and Y (in a metric or topological space) are called *topologically equivalent* (or *homeomorphic*) if there exists a bijective mapping $h : X \to Y$ such that h and h^{-1} are continuous.

A domain $D \subset \mathbb{C}$ is conformally equivalent to the unit disk $\mathbb{D}$ if and only if it is topologically equivalent to $\mathbb{D}$.

Note that there are no analogous statements for domains which are not simply connected. For example the two ring domains

$$R_1 := \{z \in \mathbb{C} : \ 1 < |z| < 2\}, \qquad R_2 := \{z \in \mathbb{C} : \ 1 < |z| < 3\},$$

are topologically equivalent, but not conformally equivalent. We shall not study this any further but refer to Nehari's classical book [43]. For a contemporary treatment see for instance Section 6.6 in Volume 2 of Lin [35].

Conformal Mappings on the Sphere. Let us finally briefly comment on conformal mappings between simply connected domains on the Riemann sphere. First of all we remark that any two of the three domains – the sphere $\widehat{\mathbb{C}}$, the plane $\mathbb{C}$, and the unit disk $\mathbb{D}$ – are not even topologically equivalent. In fact these *canonical domains* are representatives of the three equivalence classes of simply connected domains on the sphere with respect to conformal equivalence.

Theorem 6.4.2. *Any simply connected domain G on the sphere is conformally equivalent either to $\widehat{\mathbb{C}}$, $\mathbb{C}$ or $\mathbb{D}$.*

Proof. If $G = \widehat{\mathbb{C}}$ there is nothing to prove. If $G \neq \widehat{\mathbb{C}}$, we pick a point $a \in \widehat{\mathbb{C}} \setminus G$, and rotate the sphere such that a is sent to infinity. Since the image $\widetilde{G}$ of G is contained in $\mathbb{C}$ and is simply connected as a subset of $\widehat{\mathbb{C}}$, it is a simply connected domain in $\mathbb{C}$. If $\widetilde{G} = \mathbb{C}$, then G is conformally equivalent to $\mathbb{C}$, otherwise $\widetilde{G}$ satisfies the assumptions of the Riemann mapping theorem, so that $\widetilde{G}$, and hence G, is conformally equivalent to $\mathbb{D}$. $\square$

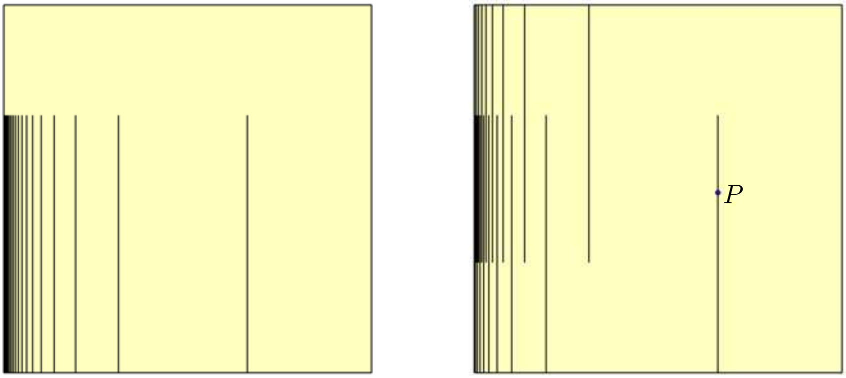


Figure 6.31: Two simply connected domains with complicated boundaries

Constructing Functions Via Conformal Mappings. Before we end this section we would like to point out that the Riemann mapping theorem is a powerful *constructive* tool: it allows one to *define* analytic functions as conformal mappings between specific domains with prescribed geometry. So these functions are not given by some explicit formula or an algorithm, but they are characterized by their geometric mapping properties. In particular, many functions with 'wild' behavior near the boundary of their domain can be described conveniently in this way. For example, you may think about a (normalized) conformal mapping of the unit disk onto the Koch snowflake domain shown in Figure 2.39, or onto one of the two domains which result from slitting a square along infinitely many segments as illustrated in Figure 6.31.

6.5 Boundary Correspondence

The definition of conformality involves only interior points of the domains and says nothing about the behavior of the mapping on the boundary – in general it is not even defined there. On the other hand, in many examples we have seen so far, the conformal mapping was the restriction of an analytic function from a larger domain, and was conformal at boundary points as well. So the question arises in which way this nice behavior is related to the geometry of the domains.

Before we formulate a central result on the boundary behavior of conformal mappings, we mention that there are many delicate problems in this field. Though the fundamental questions have been completely settled for a long time, some aspects are not yet fully understood even today and are subjects of ongoing research. Recommended reading for those who wish do delve deeper into this topic is Pommerenke's book [54].

Here we address only one single problem: which geometric conditions guarantee that a conformal mapping $f : G \to D$ can be extended continuously to a bijection between the closures $\overline{G}$ and $\overline{D}$ of the domains? The question has a completely natural answer.

Theorem 6.5.1 (Carathéodory-Osgood)**.** *If G and D are Jordan domains, any conformal mapping $f : G \to D$ can be extended to a homeomorphism between $\overline{G}$ and $\overline{D}$.*

Despite the beauty and simplicity of the result, no really easy proof is known. So we cheat a little and delegate some of the geometric difficulties encountered to the next lemma which we shall not prove. We only mention that the assertions of this lemma are obvious when G is a disk, and that the general situation can be reduced to this special case with little effort when we take Proposition 2.7.25 for granted.

Lemma 6.5.2. *If G is a Jordan domain the following assertions hold:*

(i) *Any two points $a, b \in \overline{G}$ can be connected by a Jordan arc in $G \cup \{a, b\}$ (in the following we shall simply speak of a Jordan arc in G).*

(ii) *Let $w_0, w_0' \in \partial G$ and $b, b' \in G$, with $w_0 \neq w_0'$ and $b \neq b'$. Then there exist positive numbers ε and δ with the following property: for all points $w, w' \in G$ with $|w - w_0| < \varepsilon$ and $|w' - w_0'| < \varepsilon$ there exist Jordan arcs B and B' in G, such that B connects b with w, B' connects b' with w', and $\operatorname{dist}(B, B') > \delta$.*

The first and crucial step towards the Carathéodory-Osgood theorem is the proof of the following lemma.

Lemma 6.5.3. *Any conformal mapping f of the unit disk $\mathbb{D}$ onto a Jordan domain G has a continuous extension to $\overline{\mathbb{D}}$.*

Proof. 1. We fix some point $z_0 \in \partial\mathbb{D}$ and let $(z_k) \subset \mathbb{D}$ be a sequence converging to z_0. Then $f(z_k)$ lies in G, and by the Bolzano-Weierstrass theorem it contains a subsequence which converges to some point $w_0 \in \overline{G}$.

In fact the point w_0 must lie on the boundary of G. If this were not so, its preimage $z_0' := f^{-1}(w_0)$ would lie in G. Then we could find a small disk $U \subset G$ centered at z_0' which contains only finitely many points z_k. By the open mapping principle, $V := f(U)$ covers a neighborhood of w_0. On the other hand, it contains only a finite number of the points $f(z_k)$. This is impossible since $f(z_k) \to w_0$.

2. Let us assume, for a moment, that there were two sequences (z_k) and (z_k') converging to the same point $z_0 \in \partial\mathbb{D}$, while $f(z_k)$ and $f(z_k')$ tend to different points $w_0 \in \partial G$ and $w_0' \in \partial G$, respectively. Now comes the technical part: we shall show that this cannot happen.

2.1. We pick two distinct points b, b' in G, and choose ε and δ according to assertion (ii) of Lemma 6.5.2. Further we set $a := f^{-1}(b)$, $a' := f^{-1}(b')$, and fix r_0

with $r_0 < \min\{|a - z_0|, |a' - z_0|\}$. Finally, we choose R so large that the domain G fits into a circle with radius R, and fix a positive number ϱ with

$$\varrho < \min\left\{r_0, (1/r_0)\exp(-2\pi R^2/\delta^2)\right\}. \tag{6.29}$$

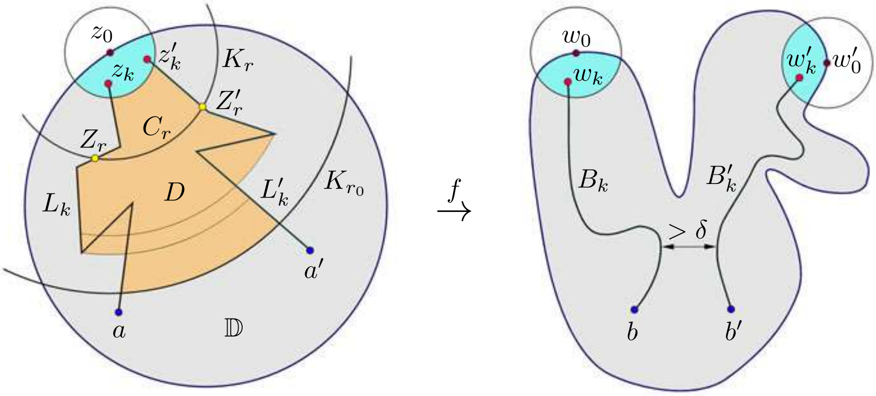


Figure 6.32: Towards a proof of the Carathéodory-Osgood theorem (schematic)

2.2. Next we set $w_k' := f/z_k'$ and choose k sufficiently large, such that $|z_k - z_0| < \varrho$, $|z_k' - z_0| < \varrho$, $|w_k - w_0| < \varepsilon$, and $|w_k' - w_0'| < \varepsilon$. Then, by Lemma 6.5.2, there exist Jordan arcs B_k, B_k' in G, connecting b with w_k and b' with w_k', respectively, such that $\operatorname{dist}(B_k, B_k') > \delta$ (see the illustration on the right in Figure 6.32).

The preimages $A_k := f^{-1}(B_k)$ and $A_k' := f^{-1}(B_k')$ are disjoint Jordan arcs in $\mathbb{D}$ which connect a with z_k and a' with z_k', respectively. These Jordan arcs can be replaced by approximating polygonal lines L_k and L_k', such that even then

$$\operatorname{dist}\left(f(L_k), f(L_k')\right) > \delta. \tag{6.30}$$

Here we have used the fact that everything plays on a compact subset of $\mathbb{D}$ where f is uniformly continuous.

2.3. For every r with $\varrho \le r \le r_0$, the circle K_r centered at z_0 with radius r intersects the polygonal lines L_k and L_k'. In the next step we choose points $Z_r \in L_k \cap K_r$ and $Z_r' \in L_k' \cap K_r$. Since K_r is a circular arc, while L_k and L_k' consist of a finite number of segments, the number of intersection points from which we can choose is finite, and we select Z_r and Z_r' such that they can be represented (in polar coordinates with center z_0) as

$$Z_r = z_0 + r\,\mathrm{e}^{\mathrm{i}\varphi(r)}, \qquad Z_r' = z_0 + r\,\mathrm{e}^{\mathrm{i}\psi(r)}, \tag{6.31}$$

where φ and ψ are piecewise continuous functions.

2.4. Let $C_r \subset \mathbb{D}$ denote the closed subarc of K_r with initial point Z_r and terminal point Z_r'. Since the images of Z_k and Z_k' lie on the curves B_k and B_k', respectively, we have (see Step 2.2)

$$0 < \delta \le |f(Z_k') - f(Z_k)| = \left| \int_{C_r} f'(z)\, dz \right| \le \int_{\varphi_-(r)}^{\varphi_+(r)} |f'(z)|\, r\, d\varphi, \tag{6.32}$$

where $\varphi_-(r) := \min\big(\varphi(r), \psi(r)\big)$, $\varphi_+(r) := \max\big(\varphi(r), \psi(r)\big)$, and $z = z_0 + r\mathrm{e}^{\mathrm{i}\varphi}$. Estimating the right-hand side of (6.32) by the Cauchy-Schwarz inequality, we get

$$\begin{aligned} \delta^2 &\le \left(\int_{\varphi_-(r)}^{\varphi_+(r)} \big(|f'(z)|\, r\big) \cdot 1\, d\varphi \right)^2 \\ &\le \left(\int_{\varphi_-(r)}^{\varphi_+(r)} |f'(z)|^2\, r^2\, d\varphi \right) \left(\int_{\varphi_-(r)}^{\varphi_+(r)} 1\, d\varphi \right) \\ &\le 2\pi \int_{\varphi_-(r)}^{\varphi_+(r)} |f'(z)|^2\, r^2\, d\varphi. \end{aligned}$$

After dividing by r and integrating with respect to r from ϱ to r_0 we arrive at

$$\delta^2 \log \frac{r_0}{\varrho} \le 2\pi \int_{\varrho}^{r_0} \int_{\varphi_-(r)}^{\varphi_+(r)} |f'(z)|^2\, r\, d\varphi\, dr. \tag{6.33}$$

The integral on the right-hand side is the area of the image $f(D)$ of the (measurable) set (see Remark 6.1.3)

$$D := \bigcup_{\varrho \le r \le r_0} C_r = \big\{ z_0 + r\,\mathrm{e}^{\mathrm{i}\varphi} : \varrho \le r \le r_0,\ \varphi_-(r) \le \varphi \le \varphi_+(r) \big\}.$$

Since $f(D)$ is contained in G, which in turn lies inside a disk with radius R, this area cannot exceed $2\pi\, R^2$. Consequently we have

$$\delta^2 \log \frac{r_0}{\varrho} \le 2\pi R^2,$$

contrary to (6.29).

3. Let $z_0 \in \partial\mathbb{D}$ and $z_k \in \mathbb{D}$ with $z_k \to z_0 \in \partial\mathbb{D}$. As was shown in the second step, the sequence $\big(f(z_k)\big)$ has exactly one accumulation point w_0, which implies that $f(z_k) \to w_0$. Since this is true, with one and the same w_0, for all sequences z_k converging to z_0, the function f can be continuously extended to z_0 by defining $f(z_0) := w_0$. The same can be done for any $z_0 \in \partial\mathbb{D}$, and then the extended function is (uniformly) continuous on the closure of $\mathbb{D}$. □

We are now fully equipped to work out a proof of the Carathéodory-Osgood theorem.

Proof. 1. Since any conformal mapping $f : D \to G$ can be represented as a composition $f = g \circ h^{-1}$, with conformal mappings $g : \mathbb{D} \to G$ and $h : \mathbb{D} \to D$, we may assume that $D = \mathbb{D}$.

2. By Lemma 6.5.3, the conformal mapping f has a continuous extension to $\overline{\mathbb{D}}$ which we again denote by f. We show that $f : \overline{\mathbb{D}} \to \overline{G}$ is *surjective.* It is clear that $f(\overline{\mathbb{D}}) \supset G$. If $w_0 \in \partial G$, there exists a sequence $(w_k) \subset G$ converging to w_0. The corresponding sequence (z_k) of preimages $z_k := f^{-1}(w_k)$ contains a subsequence which converges to some point $z_0 \in \overline{\mathbb{D}}$. Because f is continuous on $\overline{\mathbb{D}}$, we have $w_0 = f(z_0)$.

3. We show that $f : \overline{\mathbb{D}} \to \overline{G}$ is *injective.* Since $f(\mathbb{D}) = G$ and $f(\partial\mathbb{D}) = \partial G$, we only need to show injectivity of the boundary mapping.

Let $w_0 \in \partial G$. The preimage-set $A := f^{-1}(\{w_0\})$ is a closed subset of the unit circle, so that its complement $\partial\mathbb{D} \setminus A$ consists of at most countably many open arcs. Assume that there were at least two such arcs A_1 and A_2. Let L_1 be a closed segment which connects the endpoints of A_1. Then its image $J_1 := f(L_1)$ is a (closed) Jordan curve in $G \cup \{w_0\}$.

Now we pick two points a_1 in A_1 and a_2 in A_2 respectively. The segment $L := [a_1, a_2]$ intersects L_1 at exactly one point c. Since $f(a_1)$ and $f(a_2)$ lie on $\partial G \setminus \{w_0\}$, and hence in the exterior of J_1, the arcs $f\big([a_1, c)\big)$ and $f\big((c, a_2]\big)$ must be completely contained in the exterior of J_1, and thus no point of $f\big(L\big)$ can belong to the interior of J_1. On the other hand, because L and L_1 cross each other at c and f is conformal at this point, the image curves $f(L)$ and $J_1 = f(L_1)$ must intersect each other (at a non-zero angle) at $f(c)$, which is a contradiction.

Since $\partial\mathbb{D} \setminus A$ is a single open arc, A is either a point or a closed circular arc. In the latter case we have $f(z) = a = \text{const}$ on A. As we shall see in the next section (Corollary 6.6.7), this implies that f is constant in $\mathbb{D}$, which is impossible.

4. By now we know that f extends to a continuous bijective mapping of $\overline{\mathbb{D}}$ onto $\overline{G}$. Since $\overline{\mathbb{D}}$ is compact, we can invoke a standard result telling us that f^{-1} is continuous as well. □

Readers who feel uncomfortable with a proof of the Carathéodory-Osgood theorem that relies on the unproven assertions of Lemma 6.5.2 (or Proposition 2.7.25, respectively) may prefer the following reformulation:

If there exists a homeomorphism $\mathbb{D} \to G$ at all which can be extended to a homeomorphism $\overline{\mathbb{D}} \to \overline{G}$, then every conformal mapping $\mathbb{D} \to G$ will also have this property.

The Carathéodory-Osgood theorem allows an alternative normalization of conformal mappings between Jordan domains. The proof of the following corollary is left as an exercise.

Corollary 6.5.4. *Assume that D and G are Jordan domains and let (z_1, z_2, z_3) and (w_1, w_2, w_3) be equi-oriented triples of points on the boundary of D and G, respectively. Then there exists a unique homeomorphism of $\overline{D}$ onto $\overline{G}$ which maps*

D conformally onto G and satisfies

$$f(z_1) = w_1, \quad f(z_2) = w_2, \quad f(z_3) = w_3.$$

The last result of this section provides a convenient method to verify that a given function maps some domain D conformally onto another domain G. All that we need is the injectivity of the boundary mapping.

Theorem 6.5.5. *Let D be a Jordan domain and assume that $f : \overline{D} \to \mathbb{C}$ is analytic in D and continuous on $\overline{D}$. If f is injective on ∂D, then $J := f(\partial D)$ is a Jordan curve and f maps D conformally onto $G := \operatorname{int} J$.*

Proof. 1. First we reduce the problem to $D = \mathbb{D}$. Let g be a conformal mapping of $\mathbb{D}$ onto D. By the Carathéodory-Osgood theorem g extends to a homeomorphism between $\overline{\mathbb{D}}$ and $\overline{D}$, so that we may consider $f \circ g$ instead of f.

2. The image set $J := f(\mathbb{T})$ of the unit circle is a Jordan curve parameterized by $[0, 2\pi] \mapsto f\big(\mathrm{e}^{\mathrm{i}t}\big)$. For convenience we shall use the unit circle $\mathbb{T}$ as parameter space instead of the interval $[0, 2\pi]$.

If $w_0 \notin J$, the winding number $\operatorname{wind}(f, w_0)$ is defined as the winding number of the closed path $\mathbb{T} \to \mathbb{C}$, $t \mapsto f(t)$ about w_0. For all $w_0 \in \operatorname{ext} J$ we have $\operatorname{wind}(f, w_0) = 0$, and for $w_0 \in G := \operatorname{int} J$ the winding number is either equal to 1 or to -1, depending on the orientation of J.

3. For $0 < r \le 1$ we define f_r on $\mathbb{T}$ by $f_r(t) := f(rt)$. If $w_0 \notin J$, then $\operatorname{dist}(w_0, J) > 0$, and since f is uniformly continuous on $\overline{\mathbb{D}}$, we have $w_0 \notin f_r(\mathbb{T})$ and $\operatorname{wind}(f, w_0) = \operatorname{wind}(f_r, w_0)$ for all r sufficiently close to 1. By the argument principle (Theorem 3.4.5), the winding number $\operatorname{wind}(f_r, w_0)$ counts the number of solutions of $f(z) = w$ with $|z| < r$ (provided that there is no solution with $|z| = r$). Consequently, for all $w_0 \in \mathbb{C} \setminus J$, the number of solutions of $f(z) = w_0$ in $\mathbb{D}$ is equal to $\operatorname{wind}(f, w_0)$. So $f(z) = w_0$ has no solution $z \in \mathbb{D}$ if $w_0 \in \operatorname{ext} J$, and, since this number cannot be negative, there is exactly one solution $z \in \mathbb{D}$ if $w_0 \in \operatorname{int} J$.

4. In the third step we have shown that f maps $\mathbb{D}$ into $\overline{G}$. To complete the proof we must still exclude points of $\mathbb{D}$ from being mapped onto $J = \partial G$. This follows from the open mapping principle (Theorem 3.4.8): if $z_0 \in D$ with $w_0 := f(z_0) \in J$, then a complete neighborhood of w_0 is covered by $f(\mathbb{D})$. Because any such neighborhood includes points of $\operatorname{ext} J$, which are not in $f(\mathbb{D})$, this is impossible. □

The boundary behavior of conformal mappings is a classical field of research comprising many beautiful results. Among them is the *continuity theorem*, which states that a conformal mapping of $\mathbb{D}$ onto G has a continuous extension to $\overline{\mathbb{D}}$ if and only if the boundary of G is the trace of a loop (closed path).

That this result is truly remarkable becomes clear when it is rephrased as follows: if the boundary ∂G of G can be parameterized *at all* by a continuous mapping $\gamma : \mathbb{T} \to \partial G$, then this can also be accomplished by the (continuous!) boundary function of the conformal mapping $f : \mathbb{D} \to G$.

The investigation of conformal mappings of $\mathbb{D}$ onto general simply connected domains G is based on more elaborate techniques, involving concepts like *crosscuts* and *prime ends*.

Space does not permit us to introduce these notions here (interested readers are referred to Pommerenke [54], Chapter 2), we just mention that prime ends are equivalence classes which serve as substitutes for the points on the boundary of G. In the window on the right in Figure 6.31 the point P corresponds to two prime ends, while the complete left edge of the square is 'compressed' into one prime end.

The main result is that any conformal mapping $f : \mathbb{D} \to G$ establishes a natural one-to-one correspondence between the boundary points of the unit disk and the prime ends of the image domain G.

Even if a conformal mapping $f : D \to G$ has a differentiable extension to the closure of D, its extension does not need to be angle preserving at the boundary; this happens, for instance, for conformal mappings of a square onto a disk. However, appropriate additional conditions may guarantee conformality of the extended function on the boundary as well. One such situation will be considered in the next section.

6.6 The Reflection Principle

Besides Weierstrass' disk chain method, there are other principles which sometimes enable the extension of an analytic function to a larger domain. In this section we shall study one such technique which relies on *symmetry properties* and has plenty of applications also beyond conformal mappings. In particular we shall consider symmetries with respect to reflections along a line or a circle.

Definition 6.6.1. A set $D \subset \mathbb{C}$ is said to be *symmetric along the real axis*, if $z \in D$ implies that $\overline{z} \in D$.

The real line splits a set D into three disjoint subsets D_-, D_0, and D_+,

$$D_0 := \{z \in G : \operatorname{Im} z = 0\}, \qquad D_\pm := \{z \in G : \pm\operatorname{Im} z > 0\}.$$

Theorem 6.6.2 (Schwarz Reflection Principle)**.** *Let $D \subset \mathbb{C}$ be a domain which is symmetric along the real axis. Then any continuous function $f : D_+ \cup D_0 \to \mathbb{C}$ which is analytic in D_+ and real-valued on D_0 admits an analytic extension to D, namely*

$$F(z) := \begin{cases} f(z) & \text{if} \quad z \in D_+ \cup D_0 \\ \overline{f(\overline{z})} & \text{if} \quad z \in D_-. \end{cases} \tag{6.34}$$

Proof. 1. Obviously, F is analytic in D_+. In order to show that it is also analytic in D_- we pick z_0 in D_- and consider the difference quotient

$$\frac{F(z)-F(z_0)}{z-z_0} = \frac{\overline{f(\overline{z})}-\overline{f(\overline{z}_0)}}{z-z_0} = \overline{\left[\frac{f(\overline{z})-f(\overline{z}_0)}{\overline{z}-\overline{z}_0}\right]}.$$

If z tends to z_0, then $\overline{z}$ tends to $\overline{z}_0 \in D_+$. Since f is differentiable at $\overline{z}_0$, the limit of the right-hand side exists and is equal to $\overline{f'(\overline{z}_0)}$.

2. Since f is continuous in $D_+ \cup D_0$ and real-valued on D_0, the extended function F has the finite limit $f(z_0)$ at every point $z_0 \in D_0$. Consequently F is continuous in D. In order to prove that F is analytic in D, we invoke Morera's theorem (Theorem 4.2.22), which tells us that it suffices to show that the integral of F along the boundary of any triangle $\triangle$ in D vanishes. By Goursat's lemma (Lemma 4.2.16), this is certainly true if $\triangle$ is contained in D_+ or in D_-, and a continuity argument shows that it remains true if $\triangle \subset D_+ \cup D_0$ or $\triangle \subset D_- \cup D_0$. Finally, any triangle $\triangle \subset D$ can be decomposed into (at most four) sub-triangles which completely lie either in $D_+ \cup D_0$ or in $D_- \cup D_0$. The integral of f along the boundary of $\triangle$ is the sum of the integrals along the (oriented) boundaries of these sub-triangles, hence it must vanish, which completes the proof. $\square$

It follows from the identity principle that the analytic extension of f to D is unique. This yields the following corollary, which shows that symmetry properties of analytic functions in a domain D can sometimes be derived from their values on a rather small subset of D.

Corollary 6.6.3. *Let f be analytic in a domain D which is symmetric along the real axis. If f is real on $D \cap \mathbb{R}$, then $f(\overline{z}) = \overline{f(z)}$ for all $z \in D$.*

Schwarz' reflection principle has various applications. Here is a simple one about the construction of a conformal mapping involving a symmetric domain.

Example 6.6.1. We wish to construct a conformal mapping that maps the exterior E of the cross $[-1,1] \cup [-\mathrm{i},\mathrm{i}]$ onto the exterior $\mathbb{E}$ of the unit circle and preserves symmetry.

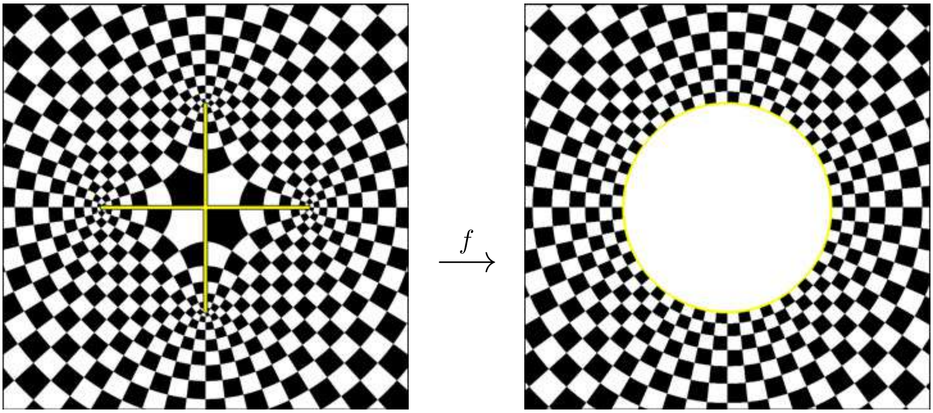


Figure 6.33: Conformal mapping of the exterior of a cross to $\mathbb{E}$

The result is visualized in Figure 6.33, the cross is depicted in yellow. Note that both domains are simply connected subsets of the Riemann sphere.

We use the fact that E and $\mathbb{E}$ are symmetric with respect to the real line. A conformal mapping g of $E \cap \mathbb{H}$ onto the *upper half-plane* $\mathbb{H}$ is already known, it is just the inverse function g of the mapping f in Example 6.2.7. The function g extends by continuity to $\mathbb{R} \setminus \{0\}$,

$$g(z) = \sqrt{z^2+1}, \qquad \operatorname{Im} z \geq 0,\ z \notin [0, \mathrm{i}],$$

where the analytic branch of the square root is chosen such that $g(z) > 0$ for $z > 0$ (which implies that $g(z) < 0$ for $z < 0$). This mapping is illustrated in the saturated part of the phase portraits in Figure 6.34.

Since the function g is real on $\mathbb{R}$, it can be extended by the Schwarz reflection principle to an analytic function G in E. The reflection is shown in brighter colors in the lower part of Figure 6.34.

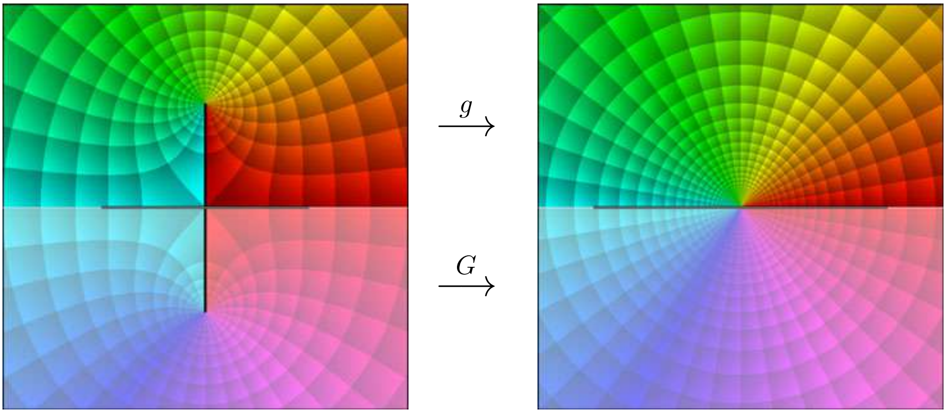


Figure 6.34: Reflection of the conformal mapping g

The relation $G(\overline{z}) = \overline{G(z)}$ guarantees that G maps E onto the complex plane slit along the segment $[-\sqrt{2}, \sqrt{2}]$. Finally, the inverse of the Joukovski map, precomposed with the dilation $z \mapsto z/\sqrt{2}$ for the right scaling,

$$F(z) := \frac{1}{\sqrt{2}}\big(z + \sqrt{z^2-2}\big), \tag{6.35}$$

sends $\mathbb{C} \setminus [-\sqrt{2},\sqrt{2}]$ to $\mathbb{E}$, so that $f := F \circ G$ is the desired conformal mapping. Its phase portrait is shown in the left window of Figure 6.35.

Combining the formulas for F and g, we obtain a nice symmetric expression,

$$f(z) = \frac{1}{\sqrt{2}}\big(\sqrt{z^2+1} + \sqrt{z^2-1}\big). \tag{6.36}$$

This is indeed a correct representation for the conformal mapping f, provided we choose the right branches of the square roots. The middle window of Figure 6.35 shows what happens when the principal branch is taken for both roots in (6.36): we get an artificial jump of the function across the imaginary axis.

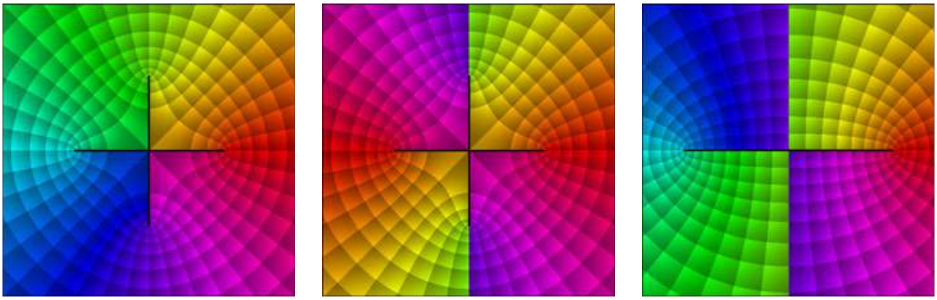


Figure 6.35: Playing with branches of the square root

We emphasize that choosing the right branch does *not* usually mean writing a plus or minus sign in front of the principal branch $\operatorname{Sqrt}(z)$; actually the sign depends on the location of z. Phase portraits often allow one to guess which sign has to be chosen in different ranges of z.

For example, the square root in the representation (6.35) of the scaled inverse Joukovski map is neither $+\operatorname{Sqrt}$ nor $-\operatorname{Sqrt}$, which (with a plus) would result in the phase portrait shown on the right in Figure 6.35. The branch cut along the segment $[-\sqrt{2}, \sqrt{2}]$ is intrinsic, but the jump across the imaginary axis is artificial. So we conjecture that the sign depends on the real part of z. Indeed, for all z with $\operatorname{Re} z \neq 0$, the correct value of the square root in (6.35) is

$$\sqrt{z^2-2} := \operatorname{sign}\big(\operatorname{Re}(z)\big) \cdot \operatorname{Sqrt}\left(z^2-2\right). \tag{6.37}$$

Interpreting the square roots in (6.36) as in (6.37) we also obtain a valid representation of the conformal mapping f.

The next result shows that appropriately normalized conformal mappings reflect the symmetry of the domains. As a by-product we obtain that simply connected mirror-symmetric domains are divided by their symmetry axes into two simply connected components.

Theorem 6.6.4. *Let D be a simply connected domain which is symmetric about the real axis, and let z_0 be a point in $D_0 := D \cap \mathbb{R}$. If $f : D \to \mathbb{D}$ is a conformal mapping with $f(z_0) \in \mathbb{R}$ and $f'(z_0) > 0$, then $f(\overline{z}) = \overline{f(z)}$ for all $z \in D$. Moreover, D_+ and D_- are simply connected domains and f maps D_+ onto the upper half circle $\mathbb{D}_+$ and D_- onto the lower half circle $\mathbb{D}_-$.*

Proof. 1. The reflected function g defined in D by $g(z) := \overline{f(\overline{z})}$ is analytic in D, it maps D conformally onto $\mathbb{D}$, and satisfies $g(z_0) = f(z_0)$, $g'(z_0) > 0$. Then the uniqueness statement of the Riemann mapping theorem implies that $f = g$, which is equivalent to $f(\overline{z}) = \overline{f(z)}$.

2. We show that the image $f(G)$ of any connected component G of $D \setminus \mathbb{R}$ is completely contained either in $\mathbb{D}_+$ or in $\mathbb{D}_-$. Assume there are two points a and b in G such that $\operatorname{Im} f(a) > 0$ and $\operatorname{Im} f(b) < 0$. When we connect a and b by a path in G, the intermediate value theorem tells us that there must exist $c \in G$ with $f(c) \in \mathbb{R}$. But then $f(\overline{c}) = \overline{f(c)} = f(c)$. Since $c \neq \overline{c} \in D$, this contradicts the injectivity of f.

3. We show that the intersection $D_0 := D \cap \mathbb{R}$ is an open (bounded or unbounded) interval. Assuming the contrary, we can find points $a, b, c \in \mathbb{R}$ with $a < b < c$ such that $a, c \in D_0$ and $b \notin D_0$. Since D is connected, there is a path $\gamma : [0,1] \to D$ from a to c in D. The point b lies on the real line in between a and c, and hence the increment of a continuous branch of the argument $t \mapsto \arg\big(\gamma(t) - b\big)$ along γ is an odd multiple of π, say $(2k+1)\,\pi$. The reflected path $\overline{\gamma}$ also lies in D and the increment of $t \mapsto \arg\big(\overline{\gamma}(t) - b\big)$ along $\overline{\gamma}$ equals $-(2k+1)\,\pi$. Now $\gamma \ominus \overline{\gamma}$ is a closed path in D with winding number $4k\pi + 2\pi \neq 0$ about $b \notin D$. This is impossible because D is simply connected.

4. Since f is real on the interval D_0, the same holds for its derivative f'. Because $f'(a) > 0$ and $f' \neq 0$, we have $f' > 0$ on all of D_0. This implies that all points of D in the upper half-plane which are sufficiently close to D_0 are mapped into $\mathbb{D}_+$. Since the closure of any component G of $D \setminus \mathbb{R}$ contains points on the real line, D_+ must be mapped into $\mathbb{D}_+$. Similarly, points in D_- are mapped into $\mathbb{D}_-$, and by the first step $f(D_0) \subset \mathbb{D}_0 := \mathbb{D} \cap \mathbb{R} = (-1, 1)$.

Since $f : D \to \mathbb{D}$ is surjective, the inclusions $f(D_-) \subset \mathbb{D}_-$, $f(D_0) \subset \mathbb{D}_0$, and $f(D_+) \subset \mathbb{D}_+$ imply that $D_- = f^{-1}(\mathbb{D}_-)$, $D_0 = f^{-1}(\mathbb{D}_0)$, and $D_+ = f^{-1}(\mathbb{D}_+)$. ☐

The Schwarz reflection principle has far-reaching generalizations. In the following version we consider reflections at circles on the Riemann sphere.

Definition 6.6.5. The *reflection R_C in a circle $C \subset \widehat{\mathbb{C}}$* is the transformation

$$R_C : \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}, \ z \mapsto \big(M_C^{-1} \circ R \circ M_C\big)(z),$$

where $R(z) := \overline{z}$ denotes the reflection across the real axis, and M_C is (any) Möbius transformation which maps C onto $\widehat{\mathbb{R}}$. A set G is said to be *symmetric about the circle C* if $R_C(G) = G$.

This definition of R_C does not depend on the choice of the Möbius transformation M_C. If C is a circle in $\mathbb{C}$ with center a and radius r the reflection R_C has a simpler description,

$$R_C(z) = a + \frac{r^2}{\overline{z} - \overline{a}},$$

which shows that the point $R_C(z)$ lies on the ray from a through z at distance $r^2/|z-a|$ from a.

Since the following theorem involves analytic functions on (subsets of) $\widehat{\mathbb{C}}$, the notion of analyticity is to be understood in the sense of Definition 3.5.1.

Theorem 6.6.6. *Let C and K be circles on $\widehat{\mathbb{C}}$ and assume that the domain $D \subset \widehat{\mathbb{C}}$ is symmetric about C. Let $D_0 := D \cap C$ and denote by D_- (respectively D_+) the union of the components of $D \setminus C$ which belong to the left (respectively right) side of C (see page 275). If a continuous function $f : D_+ \cup D_0 \to \widehat{\mathbb{C}}$ is analytic in D_0 and maps D_0 onto a subset of the circle K, then f has a unique analytic extension to D, namely*

$$F(z) := \left\{ \begin{array}{cl} f(z) & \text{if} \quad z \in D_+ \cup D_0 \\ \big(R_K \circ f \circ R_C\big)(z) & \text{if} \quad z \in D_-. \end{array} \right. \tag{6.38}$$

Proof. We only sketch the main idea. Let M_C and M_K be Möbius transformations which map C and K, respectively, to $\widehat{\mathbb{R}}$. The domain $M_C(D)$ is symmetric about the real axis. The function $g := M_K \circ f \circ M_C^{-1}$, defined on $M_C(D_+ \cup D_0)$, satisfies the assumptions of Theorem 6.6.2. (This is not the full truth, because the arguments z and the values $g(z)$ may be infinite, but these cases require only minor modifications.) If G denotes the analytic extension of g to $M_C(D)$, then $F := M_K^{-1} \circ G \circ M_C$ is the analytic extension of f across C. □

The function F will be called the *reflection of f across the circle C*. Figure 6.36 shows the process of reflecting a function across a circular arc (where it is real-valued) using the Möbius transformations M_C and M_C^{-1}.

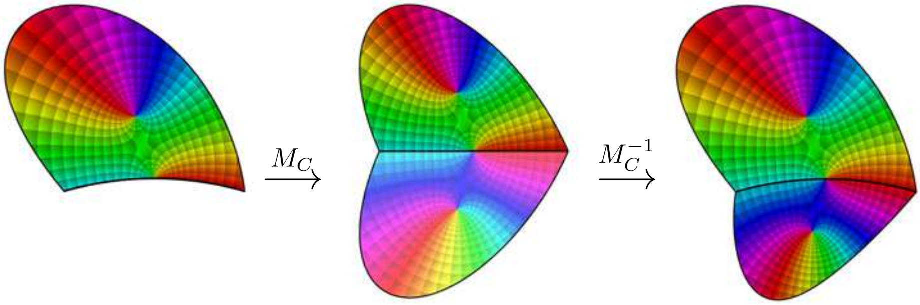


Figure 6.36: Reflecting a function which is real on a circular arc C

The next corollary is a boundary version of the identity theorem for analytic functions.

Corollary 6.6.7. *Let D be a Jordan domain whose boundary contains a Jordan arc C. If $f : D \cup C \to \widehat{\mathbb{C}}$ is continuous on $D \cup C$, analytic in D, and constant on C, then f is constant in D.*

Proof. Let g be a conformal mapping of $\mathbb{D}$ onto D. By the Carathéodory Osgood theorem, g extends to a homeomorphism between $\overline{\mathbb{D}}$ and $\overline{D}$. The function $f \circ g$ is continuous on $\overline{\mathbb{D}}$, analytic in $\mathbb{D}$, and constant on the circular arc $f^{-1}(C)$. Using the reflection principle and the identity theorem for analytic functions we conclude that $f \circ g$ is constant on $\mathbb{D}$, so that f is constant on D. $\square$

As a second corollary we obtain a result which is probably the most important application of Schwarz' reflection principle to conformal mappings.

Corollary 6.6.8. *Let $D_+ \subset \widehat{\mathbb{C}}$ be a Jordan domain and assume that its boundary contains an open circular arc D_0. Denote by D_- the reflection of D_+ about the circle C containing D_0, and suppose that $D_- \cap D_+ = \emptyset$. Assume further that f maps D_+ conformally onto a (simply connected) domain $G_+ \subset \widehat{\mathbb{C}}$, and sends D_0 onto a circular arc $G_0 := f(D_0)$. Moreover, let G_- be the reflection of G_+ along the circle K containing G_0, and assume that $G_- \cap G_+ = \emptyset$. Then f has an analytic extension F to $D := D_+ \cup D_0 \cup D_-$ which maps D conformally onto $G := G_+ \cup G_0 \cup G_-$.*

Proof. Because D_- and D_+ are Jordan domains and D_0 is an open arc which belongs to the boundaries of both domains, the set D is a domain. The Carathéodory-Osgood theorem guarantees that f is continuous on $D_+ \cup D_0$, so that it satisfies the assumptions of Theorem 6.6.6. Its extension F defined by (6.38) is analytic in D and maps D_+ and D_- conformally onto G_+ and G_-, respectively.

Invoking the Carathéodory–Osgood theorem again, we see that F is injective on $D_+ \cup D_0$, and hence on $D_- \cup D_0$. Since $f(D_-) \cap f(D_+) = \emptyset$ by assumption, F is injective on D, and finally $F(D) = G$ implies that F is a conformal mapping of D onto G. $\square$

6.7 Elliptic Integrals

Let R be a rational function in two variables, i.e., $R(x,y) = P(x,y)/Q(x,y)$ where P and Q are polynomials in x and y. It is well known from real analysis, that then the indefinite integrals

$$\int R\big(x,\sqrt{ax+b}\big)\,dx, \qquad \int R\big(x,\sqrt{ax^2+bx+c}\big)\,dx$$

can, by appropriate substitutions, be reduced to integrals over a rational function. This is, in general, not so when the square root of a polynomial of degree higher than two is involved. For polynomials of degree three and four, the integral can still be reduced to one of the three normal forms

$$\int \frac{dz}{\sqrt{(1-z^2)(1-k^2z^2)}}, \qquad \int \frac{\sqrt{1-k^2z^2}}{\sqrt{1-z^2}}\,dz, \qquad \int \frac{dz}{(1-hz^2)\sqrt{(1-z^2)(1-k^2z^2)}}$$

of the so-called (real) *elliptic integrals*. Here we study the complex version of one of these integrals, which is of special importance in the theory of conformal mappings.

Throughout this section k will denote a real number with $0 < k < 1$. We consider the functions $z \mapsto f(z;k)$ defined by

$$f(z;k) := 1/\sqrt{(1-z^2)(1-k^2z^2)}, \qquad \operatorname{Im} z \geq 0,$$

where the continuous branch of the square root is chosen such that we have $f(0;k) = 1$. The function $z \mapsto f(z;k)$ is analytic in the upper half-plane $\mathbb{H}$ and satisfies $f(\pm 1;k) = f(\pm 1/k;k) = \infty$. Its primitive

$$F(z;k) := \int_0^z \frac{dw}{\sqrt{(1-w^2)(1-k^2w^2)}}, \qquad \operatorname{Im} z > 0, \tag{6.39}$$

is called the *elliptic integral* of the *first kind* with (elliptic) *modulus* k. Figure 6.37 shows phase portraits of the functions $f(.,0.5)$ (left) and $F(.,0.5)$ (right) in the range $|\operatorname{Re} z| < 3$, $0 < \operatorname{Im} z < 6$. The dots on the lower boundary indicate the location of the points $-1/k$, -1, 0, 1 and $1/k$.

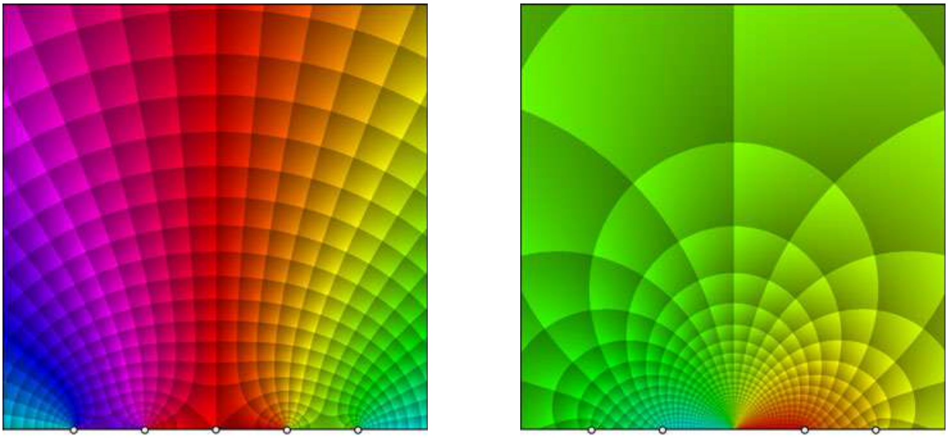


Figure 6.37: Phase portraits of $f(.,0.5)$ and the elliptic integral $F(.,0.5)$

For every $k \in (0,1)$, the elliptic integral $z \mapsto F(z;k)$ is analytic in $\mathbb{H}$ and has a continuous extension to the real line which we again denote by F. To simplify notation, we shall often make no explicit mention of the modulus k.

The mapping properties of F can be explored best when we start at the origin, where F has a zero. When z moves along the positive real axis from 0 to 1, the function f is positive so that F is real and increases monotonously. As z approaches the point $z = 1$, the value of F tends to the finite limit

$$K(k) := \int_0^1 \frac{dx}{\sqrt{(1-x^2)(1-k^2x^2)}}, \tag{6.40}$$

which is called the *complete elliptic integral* of the first kind with modulus k. When z moves further along the real axis, the value of $g(z) := (1-z^2)(1-k^2z^2)$ becomes negative. Since the branch of the square root was chosen such that f is analytic in the upper half-plane, we walk around the point 1 counterclockwise along a small semi-circle and observe what happens: the argument of $g(z)$ *decreases* from 0 to $-\pi$, so that the argument of f *increases* from 0 to $\pi/2$. Therefore the point $f(z)$ moves in the direction of the positive imaginary axis as long as $1 < z < 1/k$. When z runs through the interval $(1, 1/k)$, the increment of $F(z)$ is equal to $\mathrm{i}\,K'$, where

$$K'(k) := \int_1^{1/k} \frac{dx}{\sqrt{(x^2-1)(1-k^2x^2)}} = K\big(\sqrt{1-k^2}\big), \tag{6.41}$$

so that $F(z)$ travels through the segment $[K, K+\mathrm{i}\,K']$. When z passes the point $1/k$, the second factor $(1-k^2z^2)$ of g changes sign, the point $F(z)$ makes another turn by $\pi/2$ and starts moving in the direction of the negative real axis. If $z > 1/k$ the values $f(z)$ are *negative*, so that

$$F(z;k) = K + \mathrm{i}\,K' - \int_{1/k}^{z} \frac{dx}{\sqrt{(x^2-1)(k^2x^2-1)}}, \qquad z \in \mathbb{R},\ z \ge 1/k,$$

where the square root denotes the principal branch. As $z \to +\infty$, the integral converges, and the substitution $t := 1/(kx)$ shows that

$$\int_{1/k}^{\infty} \frac{dx}{\sqrt{(x^2-1)(k^2x^2-1)}} = \int_0^1 \frac{dt}{\sqrt{(1-k^2t^2)(1-t^2)}} = K(k).$$

Consequently, when z starts at $1/k$ and approaches $+\infty$, then $F(z)$ moves along the segment $[K+\mathrm{i}\,K', \mathrm{i}\,K']$ from $K+\mathrm{i}\,K'$ to $\mathrm{i}\,K'$.

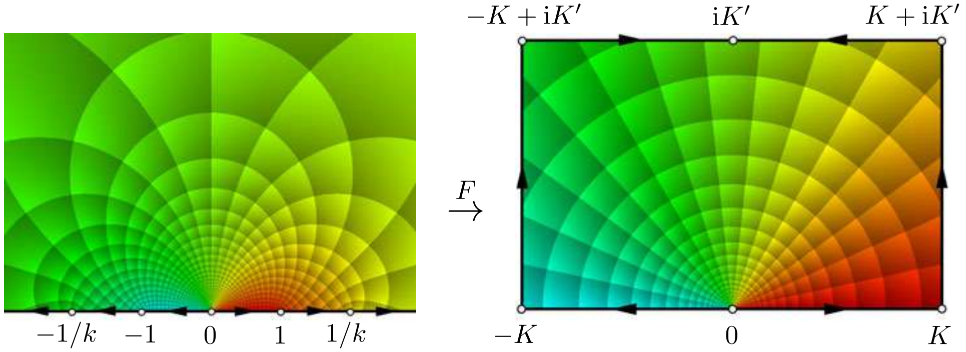


Figure 6.38: The function $F(.,0.5)$ mapping $\mathbb{R}$ to the boundary of a rectangle

Similarly we investigate what happens when z travels along the negative real axis from 0 to $-\infty$: F maps the segments $[0,-1]$, $[-1,-1/k]$ and $[-1/k,-\infty]$ onto $[0,-K]$, $[-K,-K+\mathrm{i}\,K']$ and $[-K+\mathrm{i}\,K', \mathrm{i}\,K']$, respectively.

Summarizing our observations we have: when z moves along the real axis from $-\infty$ to $+\infty$, then $F(z)$ travels monotonously once around the boundary of the rectangle

$$R(k) := \{z \in \mathbb{C} : -K < \operatorname{Re} z < K, \ 0 < \operatorname{Im} z < K'\} \tag{6.42}$$

in the counterclockwise direction. Once we know this behavior of F on the boundary, we need not investigate F in the interior of the upper half-plane – everything we are interested in can be derived from Theorem 6.5.5.

Theorem 6.7.1. *The elliptic integral $F(.;k)$ maps the upper half-plane $\mathbb{H}$ conformally onto the rectangle $R(k)$ given by (6.42), where the constants K and K' are defined in (6.40) and (6.41), respectively.*

Proof. We already know that F maps the closed real line $\widehat{\mathbb{R}}$ one-to-one onto the boundary of $R(k)$. However, the assumptions of Theorem 6.5.5 do not completely fit with the situation at hand, since F is not defined on a Jordan domain. This can be fixed by pre-composing F with the inverse g of the Cayley mapping, which maps $\mathbb{D}$ conformally onto the upper half-plane $\mathbb{H}$, so that $F \circ g$ is a good candidate for applying Theorem 6.5.5.

But there is still one detail which must be checked: we require that $F \circ g$ extends continuously onto $\overline{\mathbb{D}}$. This is obvious for all points on $\overline{\mathbb{D}}$, except for $w = 1$, which is the image of the point at infinity for the Cayley mapping. So we must show that $F(z)$ has a finite limit when z tends to infinity in the upper half-plane. This follows from the Cauchy criterion, the standard integral estimate, and the fact that the integrand f of the elliptic integral (6.39) decays like $1/|z|^2$. We leave it to the reader to work out the details. $\square$

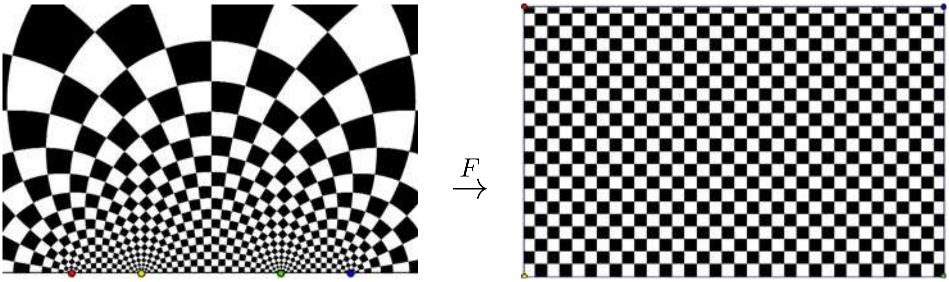


Figure 6.39: Conformal mapping of the half-plane onto a rectangle by $F(.;0.5)$

The side lengths of the image rectangle $R(k)$ are $2K(k)$ and $K'(k)$ and depend continuously on k. As $k \to 0$ we have $K(k) \to \pi/2$ and $K'(k) \to \infty$, while $K(k) \to \infty$ and $K'(k) \to \pi/2$ as $k \to 1$, so that we indeed get rectangles of arbitrary aspect ratios. The conformal mapping with $k = 0.5$ is visualized in Figure 6.39.

Mappings from the Disk. Pre-composing elliptic integrals with the inverse Cayley mapping allows us to construct conformal mappings from the disk onto rectangles. Figure 6.40 visualizes a mapping onto a square.

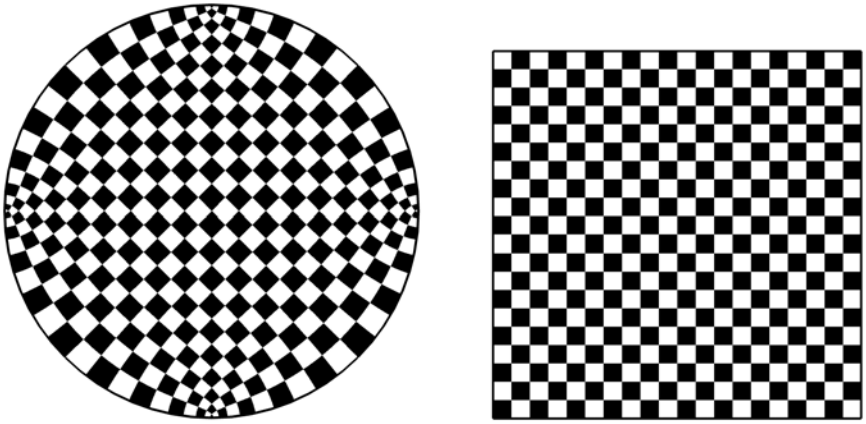


Figure 6.40: A conformal mapping from the unit disk onto a square

Jacobi Elliptic Functions. We now proceed with studying the *inverse function* $G(.;k)$ of the elliptic integral $F(.;k)$. It maps the rectangle $R_0 := R(k)$ bijectively onto the upper half-plane and has a continuous extension to the closure of R_0 as a mapping $\overline{R}_0 \to \widehat{\mathbb{C}}$. A phase portrait of G with modulus $k = 0.1535$ (where R_0 is almost a square) is shown on the left of Figure 6.41.

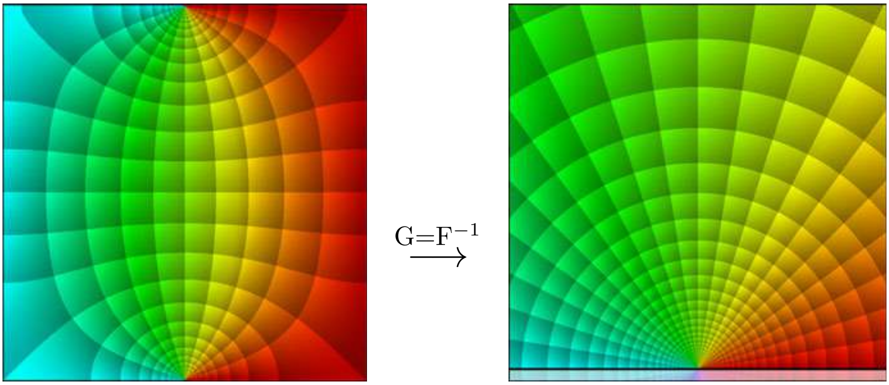


Figure 6.41: Phase portrait and mapping properties of the inverse function F^{-1}

Since the image of the segment $[K, K+\mathrm{i}\,K']$ is the interval $[1, 1/k]$ on the real axis, G can be reflected about the line L_1 through K and $K+\mathrm{i}\,K'$ (Corollary 6.6.8). The reflection of G lives on the rectangle R_1, which is the mirror image of R_0 with respect to L_1 (see the picture on the left in Figure 6.42). The reflected function maps R_1 onto the lower half-plane and satisfies the conditions for reflection along the line L_2 through $3K$ and $3K + \mathrm{i}\,K'$. This allows us to extend G further to the rectangle R_2 which is symmetric to R_1 with respect to L_2.

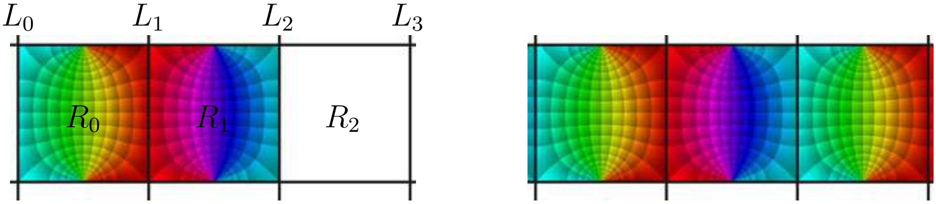


Figure 6.42: Successive reflection along vertical lines L_1, L_2, ...

The image of a point $z \in R_0$ after reflection at L_1 is $2K - z$, and after the (second) reflection at L_2 we arrive at $z + 4K$. Consequently we have

$$G(z + 4K) = \overline{G(2K - z)} = \overline{\overline{G(z)}} = G(z), \qquad z \in R_0, \tag{6.43}$$

which shows that the composition of the two reflections just acts like a *translation* of G from R_0 to R_2.

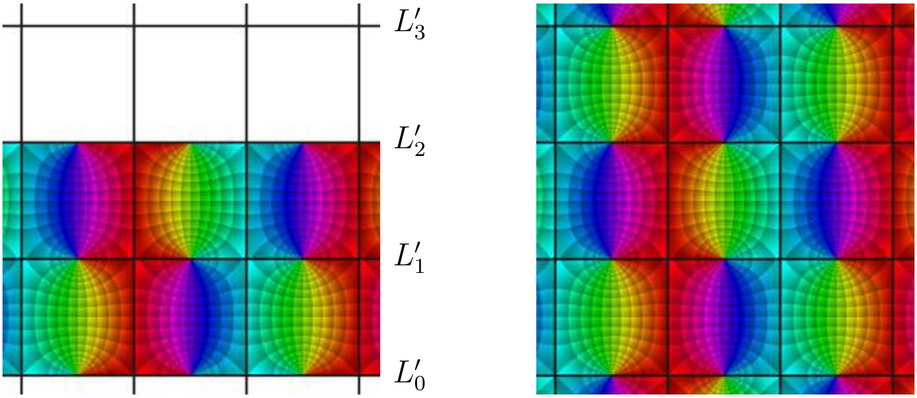


Figure 6.43: Successive reflection along horizontal lines L'_1, L'_2, ...

Similarly, G can be extended in the vertical direction from R_0 by reflection along the line L'_1 through $-K + \mathrm{i}\,K'$ and $K + \mathrm{i}\,K'$ to a rectangle R'_1. Since $G(\mathrm{i}K') = \infty$, we must apply the general form of the reflection principle (Theorem 6.6.6). The extended function G is meromorphic with a simple pole at $\mathrm{i}\,K'$.

A second reflection along the lines L'_2 through $-K + 2K'\mathrm{i}$ and $K + 2K'\mathrm{i}$ extends G to R'_2 such that

$$G(z + 2K\mathrm{i}) = G(z), \qquad z \in R_0. \tag{6.44}$$

Reflecting G successively in the horizontal direction along the straight lines L_n with $\operatorname{Re} z = (2n-1)K$ and in the vertical direction along the lines L'_n with $\operatorname{Im} z = nK'$ (for $n = 1, 2, \ldots$ and $n = -1, -2, \ldots$), we end up with a meromorphic function in the plane $\mathbb{C}$. This function is *Jacobi's elliptic sine function*, *sinus amplitudinis*, usually denoted by $\operatorname{sn} z$ or, more precisely, by $\operatorname{sn}(z; k)$. The enhanced phase portrait in Figure 6.44 depicts $\operatorname{sn}(z; 0.8)$ in the square $|\operatorname{Re} z| < 7$, $|\operatorname{Im}| z < 7$.

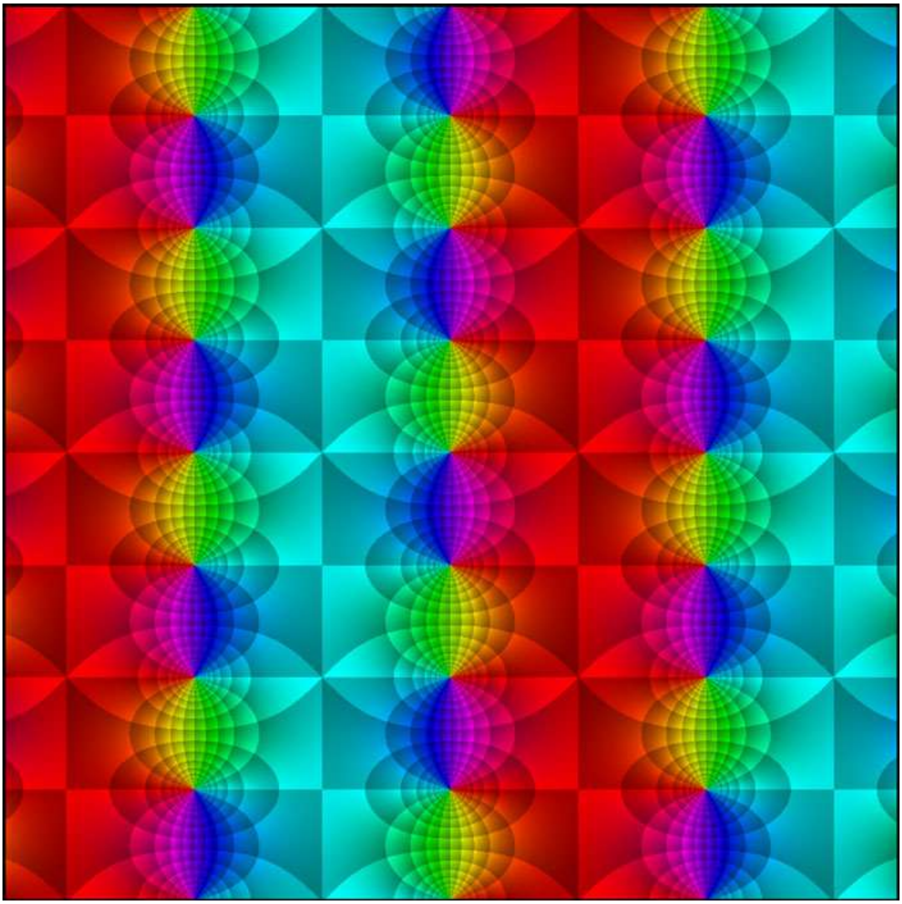


Figure 6.44: Jacobi's elliptic sine function with elliptic modulus $k = 0.8$

The functions $\operatorname{sn}(z; k)$ have simple zeros at the points $2mK + 2nK'\,\mathrm{i}$ and simple poles at $2mK + (2n+1)K'\,\mathrm{i}$ with $m, n \in \mathbb{Z}$. Most remarkable is their *double periodicity* with (primitive) periods $4K$ and $2K'\,\mathrm{i}$, which follows from (6.43) and (6.44),

$$\operatorname{sn}(z + 4K; k) = \operatorname{sn}(z; k), \qquad \operatorname{sn}(z + 2\mathrm{i}K'; k) = \operatorname{sn}(z; k), \qquad z \in \mathbb{C}. \tag{6.45}$$

The elliptic sine sn is only one out of the twelve doubly periodic Jacobi elliptic functions. Traditionally every function is denoted by an ordered pair of two letters from the set $\{\mathtt{s}, \mathtt{n}, \mathtt{d}, \mathtt{c}\}$, depending on relations between their periods and the location of zeros and poles. The three most popular functions are sn, (*sinus amplitudinis*), cn (*cosinus amplitudinis*), and dn (*delta amplitudinis*). Figure 6.45 shows phase portraits of $\mathrm{cn}\,(z; 0.8)$ and $\mathrm{dn}\,(z; 0.8)$ in the range $|\mathrm{Re}\, z| < 7$, $|\mathrm{Im}|\, z < 7$.

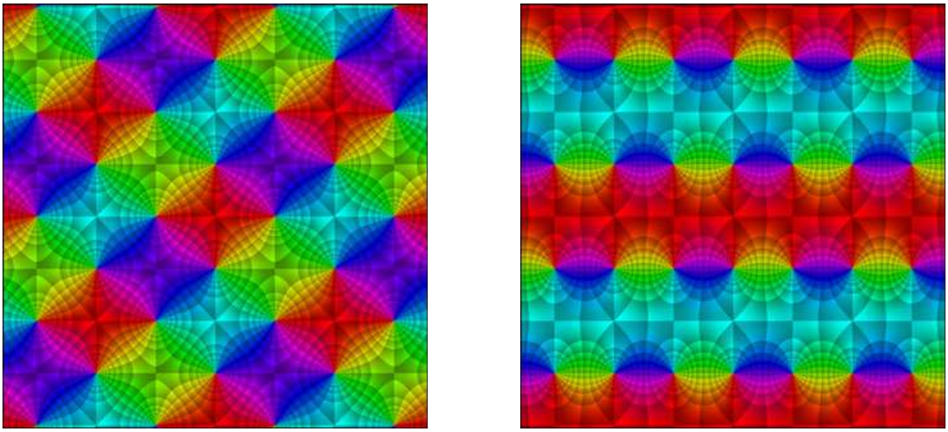


Figure 6.45: Jacobis elliptic functions cn and dn with modulus $k = 0.8$

When working with elliptic functions one should be aware that there are many different ways of denoting them. For example, instead of the *modulus* k one often finds the *parameter* m, or the *modular angle* α, which are related by

$$m = k^2 = \sin^2 \alpha.$$

In contemporary terminology, all doubly periodic meromorphic functions in the complex plane are called elliptic. We shall study this class in Volume 2.

6.8 The Schwarz-Christoffel Formula

Now that we know how to map the upper half-plane $\mathbb{H}$ conformally onto a rectangle, we are prepared for the more demanding task of mapping $\mathbb{H}$ onto an arbitrary polygon.

By a *polygon* D we understand a Jordan domain in the plane which is bounded by a *polygonal line*, that is the union of finitely many segments $[a_k, a_{k+1}]$ (with $k = 1, \ldots, n$ and $a_{n+1} = a_1$). The segments $[a_k, a_{k+1}]$ are said to be the *sides* of the polygon. We further assume that the path

$$\gamma := [a_1, a_2] \oplus [a_2, a_3] \oplus \ldots \oplus [a_{n-1}, a_n] \oplus [a_n, a_1]$$

is simple, and that the *vertices* $a_1, a_2, \ldots, a_n$ are labelled such that γ is positively oriented. Speaking of vertices, we do not require that all a_k are vertices of the polygon in the usual geometric meaning of 'corners', though (in the present setting) it does not really make sense to admit vertices on straight parts of the boundary.

The Riemann mapping theorem guarantees that any polygon D can be mapped conformally onto the upper half-plane $\mathbb{H}$, and the Carathéodory-Osgood theorem tells us that any such mapping extends to a homeomorphism between $\overline{D}$ and the closed upper half-plane $\widehat{\mathbb{H}}$. Here, as usual, the hat denotes closure with respect to the topology of the Riemann sphere.

Our aim is to derive a (more or less) explicit formula for the inverse conformal mapping $f : \widehat{\mathbb{H}} \to \overline{D}$. To get an idea what a representation of f may look like, we adapt the approach of the last section: moving the point z monotonously along the boundary of $\mathbb{H}$ (the real line) we observe what happens with the image point $f(z)$ in the w-plane. First of all we remark that $f(z)$ must move monotonously along the boundary of D, for otherwise the extension of f to $\overline{\mathbb{H}}$ could not be a homeomorphism. The key is to control the *direction* of this movement, so that $f(z)$ always stays on the boundary of the polygon D.

When a point w moves counterclockwise along the boundary of D, it only changes direction at the vertices of the polygon. If the (non-oriented) interior angle at the vertex a_k is $\alpha_k\pi$ (with $0 < \alpha_k < 2$), then w changes its direction by the (oriented) *turning* angle $\beta_k\pi$, with $\beta_k := 1 - \alpha_k$, when it passes the vertex a_k.

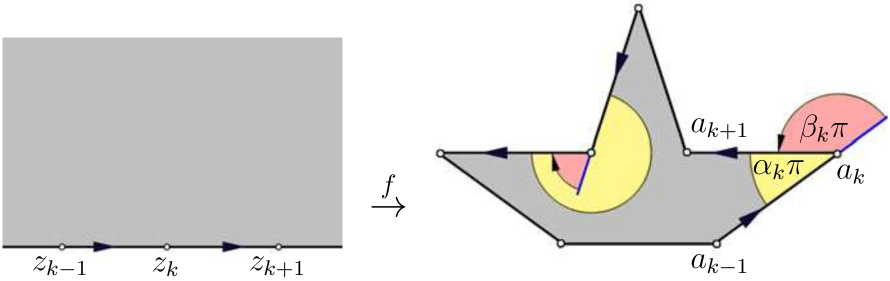


Figure 6.46: Prevertices, vertices, interior angles and turning angles of a polygon

In order to describe the movement of $f(z)$ more precisely, we apply the Schwarz reflection principle (Corollary 6.6.8). The points $z_k := f^{-1}(a_k)$ are said to be the *prevertices* of the polygon D with respect to the conformal mapping f. After applying a conformal automorphism of $\mathbb{H}$ we may assume that the prevertices satisfy $-\infty < z_1 < z_2 < \ldots < z_n < \infty$. Corollary 6.6.8 tells us that f is not only analytic in $\mathbb{H}$, but can also be analytically extended by reflection across any (open) segment (z_k, z_{k+1}). Here $(z_n, z_1) := (z_n, +\infty) \cup \{\infty\} \cup (-\infty, z_1)$ is a 'segment' on the Riemann sphere. Note that we do not speak about *one* extension here, in fact we have n *possibly different* extensions, one for every segment (z_k, z_{k+1}).

Extending f across (z_k, z_{k+1}) shows that f' has finite limits $f'(z)$ for all

$z \in (z_k, z_{k+1})$. So f' extends continuously from $\mathbb{H}$ to $\overline{\mathbb{H}} \setminus V$, where $V := \{z_1, \ldots, z_n\}$. Moreover $f'(z) \neq 0$ on $\mathbb{R} \setminus V$, for otherwise the image of a small (half-disk) neighborhood U of z in $\mathbb{H} \setminus V$ could not be contained in D because its opening angle at $f(z)$ would be at least $2\pi\mathrm{i}$.

Since f' is well-defined and different from zero on $\mathbb{R} \setminus V$, the direction in which $f(z)$ moves is determined by $\arg f'(z)$. When z runs through the interval (z_k, z_{k+1}), this direction must coincide with the direction of the corresponding image segment (a_k, a_{k+1}), that is $\arg f'$ *is constant* on (z_k, z_{k+1}). The values of these constants in neighboring segments differ precisely by the *turning angle* at their common vertex. If $\arg f'(z) = \beta_0 \pi$ for $z < z_1$, we thus have

$$\arg f'(z) = \pi\,(\beta_0 + \beta_1 + \ldots + \beta_k), \qquad z \in (z_k, z_{k+1}) \cap \mathbb{C}. \tag{6.46}$$

It is important to note that $\beta_1 + \ldots + \beta_k = 2$, which just expresses the fact that we have turned around in total by 2π, walking once around the boundary of D in counterclockwise direction. This guarantees that the prescribed values of $\arg f'(z)$ on the two components $(-\infty, z_1)$ and $(z_n, +\infty)$ of $(z_n, z_1) \cap \mathbb{C}$ coincide.

So we now have the new problem of finding analytic functions f in $\widehat{\mathbb{H}} \setminus V$, satisfying the boundary condition (6.46) on $\mathbb{R} \setminus V$. This is a typical *boundary value* problem for analytic functions. In fact we need not really solve this problem (which, as stated so far, is not even well-posed). Instead we shall use our experience from the preceding section to make a qualified guess of some *special solution*, and then show that the conformal mapping $f : \mathbb{H} \to D$ which we are looking for indeed has this form. Before we work out this in detail, let us state the result.

Theorem 6.8.1. *Let D be a polygon with vertices $a_1, \ldots, a_n$ and corresponding interior angles $\alpha_1\pi, \ldots, \alpha_n\pi$. Then there exist real numbers $z_1 < z_2 < \ldots < z_n$ and complex constants A and B such that*

$$f(z) := A \int_{\mathrm{i}}^{z} (w - z_1)^{\alpha_1 - 1} (w - z_2)^{\alpha_2 - 1} \cdots (w - z_n)^{\alpha_n - 1}\, dw + B. \tag{6.47}$$

is a conformal mapping of the upper half-plane $\mathbb{H}$ onto D.

In (6.47) the powers $(w - z_k)^{\alpha_k - 1}$ are understood in the sense of the principal branch $w^\alpha := \exp(\alpha \operatorname{Log} w)$. Since these functions can be extended analytically to the plane slit along the negative imaginary axis, this ensures that every function $f_k(z) = (w - z_k)^{\alpha_k - 1}$ is analytic in $\mathbb{H}$ and continuous on $\overline{\mathbb{H}} \setminus \{z_k\}$. In particular, every (extended) function f_k is analytic at all points of $\mathbb{R} \setminus \{z_k\}$, a fact we shall use in the following without further mention.

Proof. 1. We choose a conformal mapping f of $\mathbb{H}$ onto D which is normalized such that the prevertices $z_k := f^{-1}(a_k)$ satisfy $z_1 < z_2 < \ldots < z_n$ (see above). In order to show that f indeed has the form (6.47) — note that f is defined as a conformal mapping, and *not* by (6.47) — we study the integral term

$$f_0(z) := \int_{\mathrm{i}}^{z} (w - z_1)^{\alpha_1 - 1} (w - z_2)^{\alpha_2 - 1} \cdots (w - z_n)^{\alpha_n - 1}\, dw, \quad z \in \overline{\mathbb{H}} \setminus V. \tag{6.48}$$

This function is analytic at all points of $\overline{\mathbb{H}} \setminus V$. The argument of its derivative is constant on the segments (z_k, z_{k+1}) and jumps by $1 - \alpha_k = \beta_k$ at z_k (compare the discussion on page 297 in the preceding section). As was demonstrated above, the argument of f' has the same property, so that we get

$$\arg\, f_0'(z) - \arg\, f'(z) = c, \qquad z \in \mathbb{R} \setminus V. \tag{6.49}$$

2. Since f' and f_0' have no zeros in $\mathbb{H}$, their quotient f'/f_0 has an analytic logarithm,

$$F(z) := \log\big(f'(z)/f_0'(z)\big), \qquad z \in \mathbb{H}.$$

Because f' and f_0' also do not vanish on $\mathbb{R} \setminus V$, the function F can be extended continuously to $\mathbb{R} \setminus V$. Due to (6.49), the imaginary part $\operatorname{Im} F = \arg\, f' - \arg\, f_0'$ is constant on $\mathbb{R} \setminus V$, so that F can be reflected across any segment (z_k, z_{k+1}). In particular, the extended function (which we again denote by F) is analytic at all points of $\mathbb{R} \setminus V$. We shall show that F is also analytic at the prevertices z_k and at infinity.

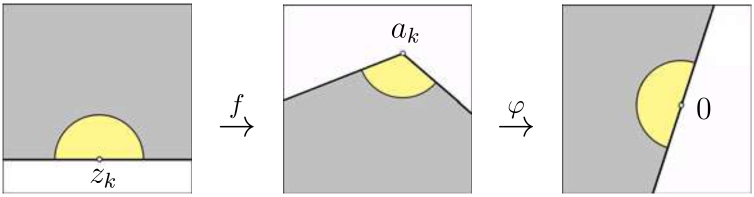


Figure 6.47: The composition $g := \varphi \circ f$ of the mappings f and φ

3. By the Carathéodory-Osgood theorem, the conformal mapping f maps a small closed semi-disk in the upper half-plane with center z_k into some sector with vertex a_k and opening angle $\alpha_k \pi$. The composition g_k of f with $\varphi_k(w) := (w - a_k)^{1/\alpha_k}$ maps this sector onto some closed half-plane (see Figure 6.47). It sends z_k to zero, and the segments $[z_{k-1}, z_k]$, $[z_k, z_{k+1}]$ are mapped to a straight line through the origin. Consequently g_k can be reflected across the segment (z_{k-1}, z_{k+1}). Since the extended function (which we again denote by g_k) is analytic at z_k and has a simple zero there, it can be represented as

$$g_k(z) = (z - z_k)\, h_k(z), \qquad h_k(z_k) \neq 0, \tag{6.50}$$

where h_k is analytic. Transforming (6.50) back to f we obtain

$$f(z) \ = a_k + (z - z_k)^{\alpha_k} \big(h_k(z)\big)^{\alpha_k}, \tag{6.51}$$
$$f'(z) = \alpha_k\, (z - z_k)^{\alpha_k - 1} \big(h_k(z)^{\alpha_k} + (z - z_k)\, h_k(z)^{\alpha_k - 1}\, h_k'(z)\big). \tag{6.52}$$

Forming the quotient f'/f_0', the factor $(z - z_k)^{\alpha_k - 1}$ cancels, so that f'/f_0' is the quotient of two analytic functions which are non-zero in a neighborhood of z_k. Consequently, f'/f_0' and $F = \log\big(f'/f_0'\big)$ are analytic at z_k.

4. In order to study the behavior of f at infinity, we consider the function g, which is defined in the lower half-plane by $g(z) := f\big(1/z\big)$. It sends 0 to the point $b := f(\infty)$ on the segment (a_n, a_1). The image of $(z_n, z_1) \subset \widehat{\mathbb{C}}$ with respect to the mapping $z \mapsto 1/z$ is an open interval I which contains 0. Since g maps I to the side (a_n, a_1) of D, it can be reflected across I. The reflected function g is analytic at 0 and its derivative does not vanish, so that it admits a representation $g(z) = b + z\,h(z)$, where h is analytic at 0 and $h(0) \neq 0$. After transforming this back to f we obtain

$$f(z) = g(1/z) = b + \frac{h(1/z)}{z}, \qquad f'(z) = -\frac{h(1/z)}{z^2} - \frac{h'(1/z)}{z^3}. \tag{6.53}$$

The behavior of f_0' at infinity is governed by the sum of the exponents $\alpha_k - 1$. Since this sum equals -2, we have

$$f_0'(z) = (z - z_1)^{\alpha_1 - 1} \cdots (z - z_n)^{\alpha_n - 1} = \frac{h_0(1/z)}{z^2}, \tag{6.54}$$

where h_0 is analytic at zero and $h_0(0) \neq 0$. Combining (6.53) and (6.54) we see that f'/f_0' is analytic and non-zero at infinity, which implies that $F = \operatorname{Log}(f'/f_0')$ is analytic at infinity as well.

5. In summary we have shown that $F : \widehat{\mathbb{H}} \to \mathbb{C}$ is analytic at all points of the closed upper half-plane $\widehat{\mathbb{H}}$. Since its imaginary part is constant on the segments (z_k, z_{k+1}), it must be constant on all of the extended real line $\widehat{\mathbb{R}}$. Then the reflection principle allows us to extend F to a bounded analytic function on the sphere. By Liouville's theorem $F = \log(f'/f_0')$ is constant, i.e., $f' = A\,f_0'$ with some $A \in \mathbb{C}$, so that $f = Af_0 + B$ for some constant B. □

So far we have assumed that all prevertices z_k are finite. If one of the z_k is the point at infinity, say $z_n = \infty$, we get the modified Schwarz-Christoffel formula

$$f(z) = A \int_{\mathrm{i}}^{z} (w - z_1)^{\alpha_1 - 1}(w - z_2)^{\alpha_2 - 1} \cdots (w - z_{n-1})^{\alpha_{n-1} - 1}\, dw. \tag{6.55}$$

Since the prevertices z_k are not known *a-priori*, the formulas (6.47) and (6.55) do not really provide an *explicit* expression for the desired conformal mapping. The determination of the prevertices is the *Schwarz-Christoffel parameter problem.* For the majority of mapping problems the parameter problem cannot be solved analytically and one has to take recourse to numerical methods. We shall briefly comment on this subject at the end of this section.

On the other hand, the parameter problem completely disappears when the polygon D is a *triangle*, since then any set of prevertices z_1, z_2, z_3 with the right ordering will do. If the mapping f of $\mathbb{H}$ onto D is normalized so that the prevertices 0, 1 and ∞ are mapped to the triangle vertices a_1, a_2, a_3 with interior angles $\alpha_1\,\pi$, $\alpha_2\,\pi$, and $\alpha_3\,\pi$, respectively, then f is given by

$$f(z) = a_1 + \frac{a_2 - a_1}{B(\alpha_1, \alpha_2)} \int_0^z w^{\alpha_1 - 1}(1 - w)^{\alpha_2 - 1}\, dw, \tag{6.56}$$

where

$$B(\alpha_1, \alpha_2) := \frac{\pi}{\Gamma(\alpha_1)\,\Gamma(\alpha_2)\,\Gamma(\alpha_3)\sin\pi\alpha_3}. \tag{6.57}$$

We remark that (6.57) with $\alpha_3 = 1 - \alpha_1 - \alpha_2$ defines *Euler's Beta-function* B. Figure 6.48 (left) shows the phase portrait of a conformal mapping of a triangle with interior angles 30°, 60°, and 90° onto the upper half-plane. The mapping is normalized so that the left, lower right, and upper right vertex of the triangle are sent to 1, ∞ and 0, respectively. The chessboard coloring in the picture on the right illustrates a mapping of the same triangle onto a circumscribed disk. This mapping is normalized so that the vertices of the triangle are fixed.

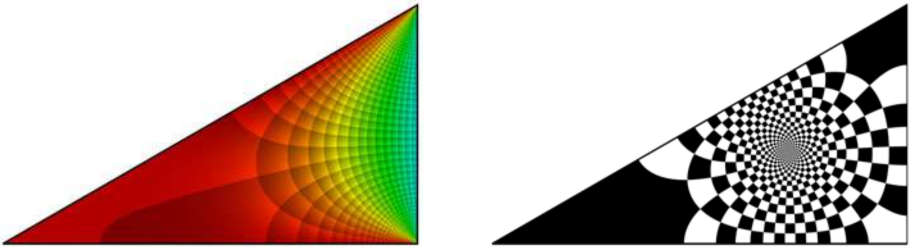


Figure 6.48: Conformal mappings of a triangle onto $\mathbb{H}$ and $\mathbb{D}$

The cases where repeated reflections of the triangle D across its sides tessellate the plane – like in the example at hand – are of special interest. We shall come back to this issue in Volume 2.

Generalizations. The Schwarz-Christoffel formula can be modified and generalized in various ways:

(1) Incorporating the formulas for standard conformal mappings from a domain D to the half-plane, we obtain representations for conformal mappings of *other canonical domains* (like disks, strips or rectangles) onto polygons.

(2) Reversing the order of the vertices and replacing interior angles by *exterior angles* of the polygon yields formulas for mappings onto *exterior domains.*

(3) One can admit that the polygon has multiple vertices (so that its boundary is not a Jordan curve), and allow angles of sizes 0 and 2π. Finally, the polygon may be self-overlapping, which yields conformal mappings which are not injective an thus can be described better in the setting of Riemann surfaces.

Recommended reading for those who would like to learn more about these aspects is the book [10] by Driscoll and Trefethen. We also refer to Lin's textbook [35], Vol. 2, which contains theoretical background as well as a number of worked-out examples.

Figure 6.49 illustrates a normalized conformal mapping of the unit disk onto a pentagram, which is given by the simple formula

$$f(z) = A \int_0^z \frac{(1+w^5)^{2/5}}{(1-w^5)^{4/5}}\, dw.$$

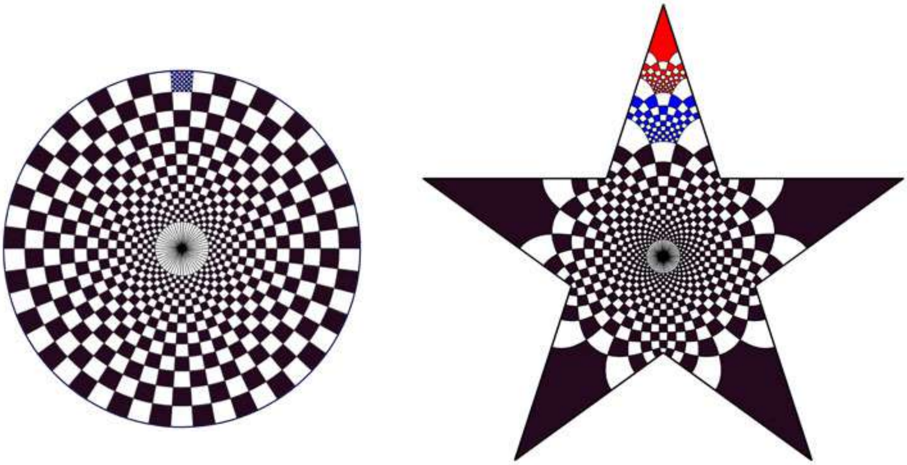


Figure 6.49: Conformal mapping of the unit disk to the interior a pentagram

Note the large distortion at the tips of the star: the barely visible tiny red domain in the disk near the point i is mapped onto the relatively large region filled with the red chessboard structure.

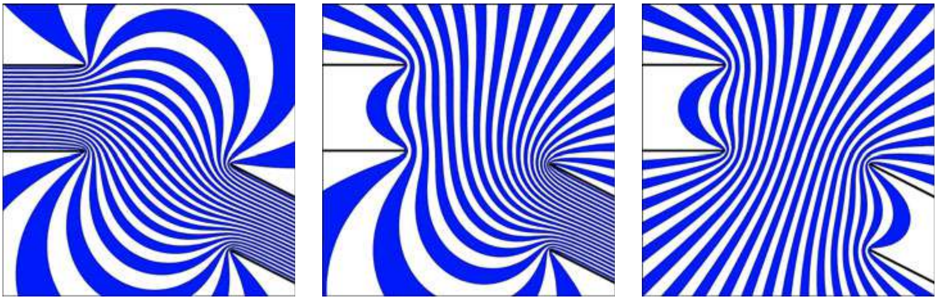


Figure 6.50: Stream lines of three different potential flows in a slit domain

As an illustration of Schwarz-Christoffel mappings onto unbounded generalized polygons, Figure 6.50 shows the stream lines of plane potential flows which are obtained from three differently normalized conformal mappings of an infinite strip to the slit domain depicted. All mappings are computed numerically using the MATLAB -based *SC Toolbox* by Driscoll and Trefethen.

The development of *numerical algorithms* for Schwarz-Christoffel mappings gathered speed with the work of Lloyd Trefethen around 1980, when Trefethen invented a reliable algorithm to numerically solve the parameter problem. But another serious obstacle remained: the so-called *crowding phenomenon*. It refers to the fact that for 'elongated domains' D some prevertices are separated by a very small distance only, which makes numerical computations problematic. The problem was resolved in the nineties by Driscoll and Vavasis, who developed a novel approach based on cross-ratios and Delaunay triangulation.

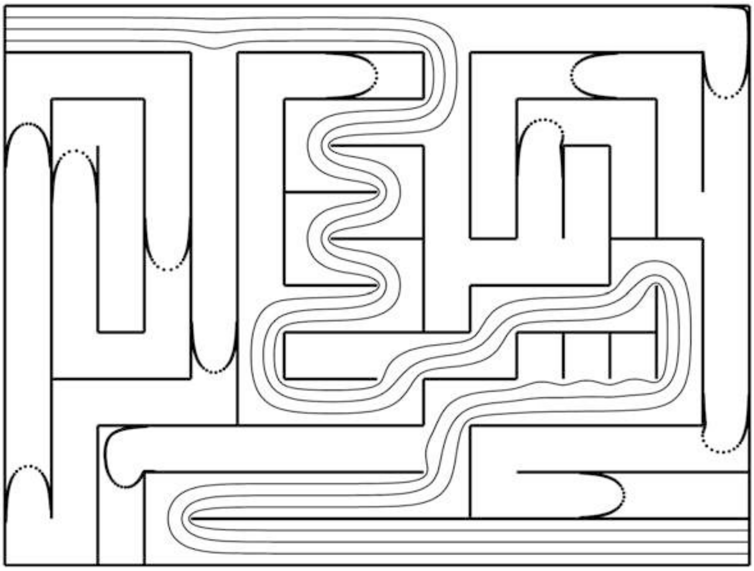


Figure 6.51: Mapping of a rectangle onto a maze computed with the SC toolbox

All these ideas have been incorporated into the extremely powerful and convenient-to-use SC Toolbox. The true capability of this software package is illustrated in Figure 6.51,[2] which visualizes the mapping of a rectangle onto a 'maze'. Each curve is the image of a line parallel to the longer sides of a (conformally equivalent) rectangle. The dotted curves correspond to lines which are at distances 10^{-10}, 10^{-20}, 10^{-30}, 10^{-40} and 10^{-50} to a rectangle side. All computations are performed in double precision arithmetic.

The algorithms implemented in the SC Toolbox are so efficient that they even allow computation of conformal mappings by approximating general domains by polygons with a large number of vertices. The book [10] by Driscoll and Trefethen and the User's Guide [11] acquaint the reader with many facets of Schwarz-Christoffel mappings and their applications, describe the basics of the underlying algorithms, and give instructions on the usage of the toolbox.

---

[2] Reproduced from [10] with permission

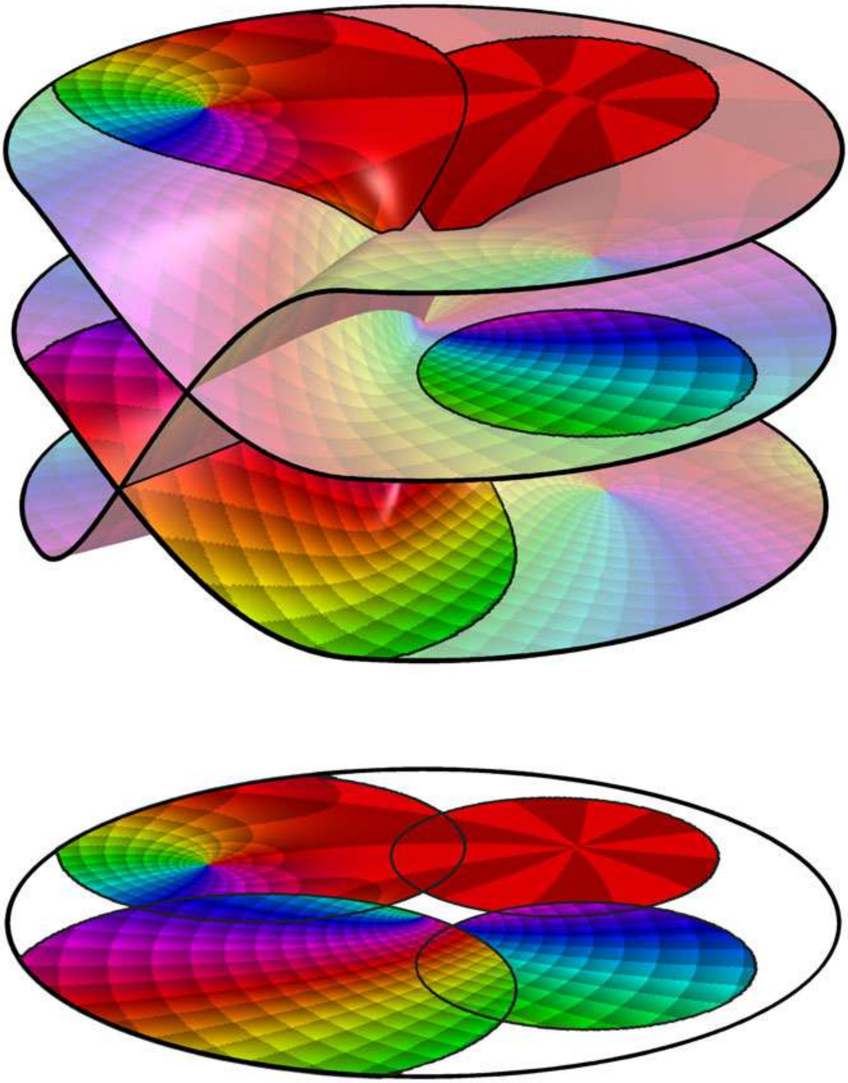


Figure 7.1: Building a Riemann surface as patchwork of function elements

Chapter 7

Riemann Surfaces

In Section 3.6 we have seen that the process of analytic continuation may result in "multiple-valued" functions. Notwithstanding their name, these objects are not ordinary functions and must be handled with care. That even simple operations involving multiple-valued functions may be problematic becomes evident when one tries to define the sum $\sqrt{z}+\sqrt{z}$ and then compares the result with $2\sqrt{z}$.

In this chapter we are going to develop an alternative framework, allowing us to avoid the drawbacks of multiple-valued functions. It is based on a geometric construction that creates a new domain, called a *Riemann surface*, comprising the different branches of the multiple-valued function. This gives rise to a single-valued analytic function on a Riemann surface. To demonstrate the basic idea we consider a simple instructive example.

Example (The Riemann Surface of the Square Root). The multiple-valued *square root* attains *two values* at every point z in the punctured plane $\mathbb{C}\setminus\{0\}$. We associate these values with *two copies* S_1 and S_2 of $\mathbb{C}\setminus\{0\}$, which we denote as *sheets*. In doing so, we encounter one major problem: there is no way to assign the values so that the resulting functions are continuous on all of $\mathbb{C}\setminus\{0\}$, so we must allow for a *line of discontinuity*. One possible, and perhaps the most natural, choice for such a line is the negative real axis $\mathbb{R}_-$.

After cutting the sheets S_1 and S_2 along $\mathbb{R}_-$, we define two analytic branches σ_1 and σ_2 of the square root, specified by $\sigma_1(1)=1$ and $\sigma_2(1)=-1$, respectively. Every branch σ_k lives on the corresponding sheet S_k (see Figure 3.42), both functions are discontinuous at the cut, but their one-sided limits exist. We emphasize that 'cutting along the negative real axis' does not mean that the segment $(-\infty,0]$ is removed, instead we restore the interval $(-\infty,0)$ to the "upper" edge (where $\operatorname{Im} z>0$) of the slit. So every sheet still contains all points of $\mathbb{C}\setminus\{0\}$, we only change the topology in a neighborhood of the cut.

When we now stack one sheet on top of the other, the two edges of S_1 can be glued to the opposite edges of S_2, so that the two branches of the square root

E. Wegert, *Visual Complex Functions: An Introduction with Phase Portraits*,
DOI 10.1007/978-3-0348-0180-5_7, © Springer Basel 2012

form a *continuous* function on the resulting geometric object S, as is illustrated in Figure 7.2. We call S the *concrete Riemann surface* of the square root.

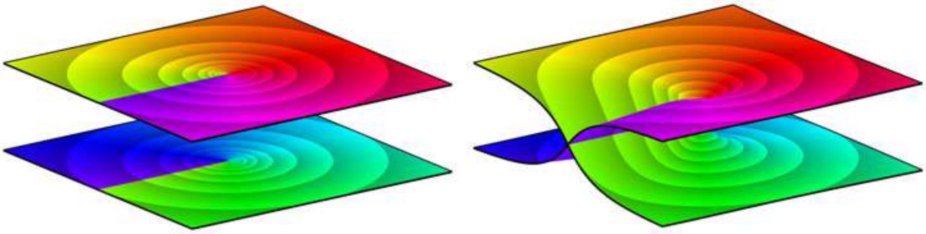


Figure 7.2: Building the concrete Riemann surface of the square root

So we have created a new domain S, on which the square root is a *well-defined* (single-valued) function. The multi-valuedness has been resolved by *lifting* the function from the plane to its Riemann surface. From this viewpoint the square root just *appeared* to have two values at a single point, because we did not distinguish between points on different sheets of its Riemann surface which have the same projection onto the plane. The Riemann surface S is therefore the *natural domain of definition* of the square root.

The above construction is not only helpful to get an intuitive understanding, it also allows one to build and explore Riemann surfaces of special functions. Yet the approach is unsatisfactory because it relies too much on visual persuasion. In this chapter we shall put these ideas on a more solid basis, working out the necessary framework step-by-step at increasing levels of abstraction.

7.1 Global Analytic Functions

Our starting point is a collection of (analytic) function elements (see Definition 3.6.2) from which we construct a model of a new geometric object – the Riemann surface of a global analytic function.

Definition 7.1.1. A *global analytic function* in a domain D is a non-empty set $\mathcal{F}$ of analytic function elements in D having the following two properties:

(i) Any two function elements belonging to $\mathcal{F}$ are analytic continuations of each other along some path in D.

(ii) Any function element which can be obtained by analytic continuation of an element in $\mathcal{F}$ along a path in D is itself a member of $\mathcal{F}$.

So a global analytic function in D is the family $\mathcal{F}$ of all analytic function elements that can be obtained from one (and then any) member (f_0, D_0) of $\mathcal{F}$ by all possible analytic continuations along paths in D starting at the center of D_0. We also say that $\mathcal{F}$ is generated by (f_0, D_0). If the domain D is not specified, we implicitly assume that $D = \mathbb{C}$. Occasionally we also allow that $D = \widehat{\mathbb{C}}$.

Instead of function elements we shall often work with the corresponding germs (see Definition 3.6.6). In the above definition function elements have been chosen as the primary objects because they have a simpler structure (they are proper functions) and can be "glued together" more intuitively (like patches) than germs. In the following we do not strictly distinguish between function elements and germs, for instance we shall also speak of a germ f^* in $\mathcal{F}$.

Example 7.1.1 (**The Global Square Root Function**)**.** We consider the totality $\mathcal{F}$ of function elements (f, D), where D is a disk in $\mathbb{C} \setminus \{0\}$, and

$$f(z) := \sqrt{|z|} \cdot \exp\left(\frac{\mathrm{i}}{2} \arg z\right), \qquad z \in D,$$

with a continuous branch $\arg z$ of the argument in D. It is not hard to see that $\mathcal{F}$ is a global analytic function.

From Global Analytic Functions to Riemann Surfaces. The sceptical reader may wonder what is gained by replacing a multiple-valued function with a large pool of function elements. In fact not much, at least for the moment, but this will change when $\mathcal{F}$ is endowed with an appropriate structure. In the first step of this construction the Riemann surface of a global analytic function will be modelled from the germs of its function elements.

Let $\mathcal{F}$ be a global analytic function. A germ which is represented by some function element (f, D) in $\mathcal{F}$ is called a germ in $\mathcal{F}$. The set of germs in $\mathcal{F}$ is denoted by $\mathcal{F}^*$, and we let $\mathcal{F}_z^*$ signify the set of all germs at z which belong to $\mathcal{F}^*$. It may happen that $\mathcal{F}_z^*$ is empty, but the next theorem shows that it can never be very large.

Theorem 7.1.2 (Poincaré-Volterra)**.** *For every global analytic function $\mathcal{F}$ and every $z \in \mathbb{C}$ the set $\mathcal{F}_z^*$ of germs in $\mathcal{F}$ at z is at most countable.*

Proof. Let (f_0, D_0) be a fixed element in $\mathcal{F}$ with center z_0. Then any element (f, D) in $\mathcal{F}$ centered at z is the analytic continuation of (f_0, D_0) along some path γ from z_0 to z. By Lemma 3.6.5 the germ f_z^* of (f, D) at z does not depend on the chain of disks $(D_0, D_1, \ldots, D_{n-1}, D)$ covering γ which are chosen in the process of analytic continuation. Moreover, if $(D_0, \widetilde{D}_1, \ldots, \widetilde{D}_{n-1}, D)$ is another chain of disks such that $\widetilde{D}_k \subset D_k$ for $k = 1, \ldots, n-1$, then the analytic continuations of (f_0, D_0) along both chains coincide. In particular, for any chain $(D_0, D_1, \ldots, D_{n-1}, D)$ covering γ, the corresponding disks $\widetilde{D}_1, \ldots, \widetilde{D}_{n-1}$ can be chosen such that their radii are rational and their centers have rational coordinates. Since the set of all chains formed from such disks is countable, the number of different germs in $\mathcal{F}$ at z is at most countable. □

In the construction of a concrete Riemann surface in Example 7, a multiple-valued function was unravelled by associating its values to different sheets 'lying above' the original domain. The number of sheets above some point z was equal to the number of values which the function attains at z. In the abstract setting, we need so many layers at every point $z \in \mathbb{C}$ as there are germs in $\mathcal{F}_z^*$, and we label the points above z by the corresponding germs. So we write $p = (z, g)$ to indicate that p is the point above z on the sheet associated with the germ g in $\mathcal{F}_z^*$. This is an intuitive interpretation – formally p is just an element of $\mathbb{C} \times \mathcal{F}^*$.

Definition 7.1.3. The *Riemann surface of the global analytic function* $\mathcal{F}$ is the set

$$S(\mathcal{F}) := \{(z,g) \in \mathbb{C} \times \mathcal{F}^* : z \in \mathbb{C},\ g \in \mathcal{F}_z^*\}. \tag{7.1}$$

The function $\mathcal{P} : S(\mathcal{F}) \to \mathbb{C},\ (z,g) \mapsto z$ is called the *(canonical) projection* of the Riemann surface $S(\mathcal{F})$ to the complex plane. The preimage set $\mathcal{P}^{-1}(z)$ of a point $z \in \mathbb{C}$ is termed the *fiber* of z (or above z). It consists of all points p on $S(\mathcal{F})$ which have been described before as 'lying above' z. More generally, we shall say that some set $A \subset S(\mathcal{F})$ lies above $B \subset \mathbb{C}$ if $\mathcal{P}(A) \subset B$.

7.2 Lifting Techniques

After these preparations we are ready to realize our main goal of incorporating all function elements of a global analytic function $\mathcal{F}$ into one single function on its Riemann surface.

Definition 7.2.1. The *lift* of a global analytic function $\mathcal{F}$ to the Riemann surface $S(\mathcal{F})$ is defined by

$$\widetilde{\mathcal{F}} : S(\mathcal{F}) \to \mathbb{C},\ (z,g) \mapsto g(z). \tag{7.2}$$

Note that this definition makes sense; because g is a germ at z it can be evaluated at the point z.

In order to explain the process of lifting better, we consider a point $p_0 = (z_0, g_0)$ on $S(\mathcal{F})$ and choose a function element (f, D_0) in $\mathcal{F}$ which represents the germ g_0. For z in D_0 the analytic function f induces a germ f_z^* at z. Since f_z^* is the analytic continuation of $f_{z_0}^* = g_0$ along the path $[z_0, z]$ in D_0, we have $(z, f_z^*) \in S(\mathcal{F})$ for all $z \in D_0$. The set

$$\widetilde{D}_0 := \{(z, f_z^*) :\ z \in D_0\} \tag{7.3}$$

is called the *lift* of the disk D_0 to p_0. We shall also refer to $\widetilde{D}_0$ as the *disk on* $S(\mathcal{F})$ *above* D_0 *with center* p_0, or as the disk on $S(\mathcal{F})$ *associated with the function element* (f, D_0) in $\mathcal{F}$. Intuitively, one can consider $\widetilde{D}_0$ as a replica of the disk D_0 on the Riemann surface $S(\mathcal{F})$. Note that $\widetilde{D}_0$ is associated with a *function element* (f, D_0) and not with the disk D_0 alone.

The restriction $\mathcal{P}_0$ of the canonical projection $\mathcal{P}$ to $\widetilde{D}_0$ is a bijection between $\widetilde{D}_0$ and D_0. The mappings

$$\mathcal{P}_0 : \widetilde{D}_0 \to D_0, \qquad \mathcal{P}_0^{-1} : D_0 \to \widetilde{D}_0,$$

are said to be a *chart* on $S(\mathcal{F})$ at p_0 and a *lifting map* to $S(\mathcal{F})$ at p_0, respectively.

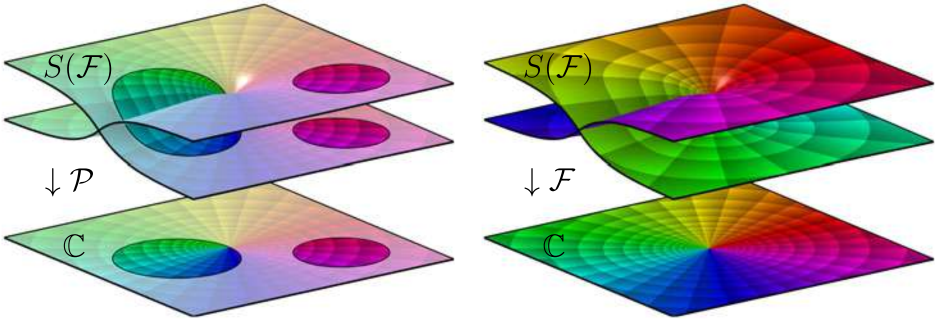


Figure 7.3: Some disks in $\mathbb{C}$ and on $S(\mathcal{F})$, the projection $\mathcal{P}$, and the function $\widetilde{\mathcal{F}}$

The restriction of $\widetilde{\mathcal{F}}$ to the disk $\widetilde{D}_0$ is given by

$$\widetilde{\mathcal{F}}(z, f_z^*) := f_z^*(z) = f(z), \qquad z \in D_0, \tag{7.4}$$

which can concisely be written as $\widetilde{\mathcal{F}} \circ \mathcal{P}_0^{-1} = f$ on D_0. In this sense the function $\mathcal{P}_0^{-1}$ 'lifts' the function elements in $\mathcal{F}$ from the plane to their associated disks on the Riemann surface. In more picturesque language, the function $\widetilde{\mathcal{F}}$ and the Riemann surface $S(\mathcal{F})$ are composed from a patchwork of the function elements in $\mathcal{F}$ as is shown in Figure 7.4 (see also the picture on page 310).

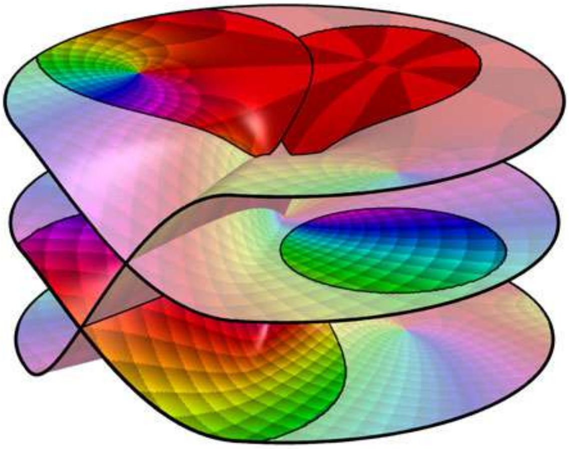


Figure 7.4: A patchwork of function elements forming a Riemann surface

The Topology of a Riemann Surface. The Riemann surface $S(\mathcal{F})$ of a global analytic function $\mathcal{F}$ inherits further structure from $\mathbb{C}$ via the canonical projection $\mathcal{P} : S(\mathcal{F}) \to \mathbb{C}$. We have already seen an instance of such a procedure when we introduced the concept of disks on $S(\mathcal{F})$. Here we utilize these disks to endow $S(\mathcal{F})$ with a topology.

If $p_0 = (z_0, g_0)$ is a point on $S(\mathcal{F})$, then g_0 is a germ at z_0 which belongs to $\mathcal{F}$. The *canonical representative* of g_0 is a function element (f, D_0), where D_0 is a disk centered at z_0 with (maximal) radius r_0 (see Definition 3.6.6). At every point $z \in D_0$ the function f induces a germ f_z^* at z. For $0 < r \leq r_0$ we denote by $D(z_0, r)$ the disk centered at z_0 with radius r and lift it to a disk $\widetilde{D}(p_0, r)$ on $S(\mathcal{F})$ with radius r and center p_0,

$$\widetilde{D}(p_0, r) := \{(z, f_z^*) : z \in D(z_0, r)\}. \tag{7.5}$$

In the language of topology, the set of disks $\widetilde{D}(p_0, r)$ with $0 < r \leq r_0$ forms a *neighborhood basis* for the point p_0 on $S(\mathcal{F})$. This endows the Riemann surface $S(\mathcal{F})$ with a *topology* and converts it into a *Hausdorff topological space.*[1]

As a consequence, all concepts of topology are now available on $S(\mathcal{F})$, such as open, closed, compact, and connected sets, as well as convergence of sequences, and continuity of functions. In particular, it is easily seen that the projection $\mathcal{P}$ and the function $\mathcal{F}$ defined by (7.2) are continuous functions from $S(\mathcal{F})$ into $\mathbb{C}$.

Path Lifting and Connectivity. As usual, a *path* on $S(\mathcal{F})$ is a continuous function $\Gamma : [\alpha, \beta] \to S(\mathcal{F})$. The *projection* of Γ to $\mathbb{C}$ is the mapping $\gamma := \mathcal{P} \circ \Gamma$. Any path on $S(\mathcal{F})$ has the form $\Gamma = (\gamma, G)$, where γ is a path in $\mathbb{C}$ and

$$G : [\alpha, \beta] \to \mathcal{F}^*, \ t \mapsto G(t) \in \mathcal{F}^*_{\gamma(t)}.$$

The continuity of the mapping Γ expresses the fact that $t \mapsto G(t)$ represents the analytic continuation of the germ $G(\alpha)$ along γ.

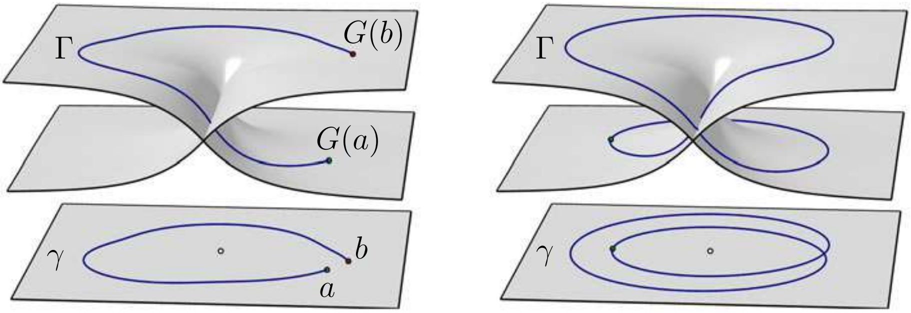


Figure 7.5: Lifting paths from the complex plane to a Riemann surface

---

[1] The attribute “Hausdorff” indicates that distinct points can be separated by disjoint neighborhoods.

Conversely, if $G : [\alpha, \beta] \to \mathcal{F}^* : t \mapsto G(t)$ is the analytic continuation of a germ $G(\alpha)$ in $\mathcal{F}^*$ along some path $\gamma : [\alpha, \beta] \to \mathbb{C}$, then

$$\Gamma : [\alpha, \beta] \to S(\mathcal{F}),\ t \mapsto \big(\gamma(t), G(t)\big) \tag{7.6}$$

is a path on $S(\mathcal{F})$. This path lifting technique also shows that $S(\mathcal{F})$ is connected: if $p_1 = (z_1, g_1)$ and $p_2 = (z_2, g_2)$ are points on $S(\mathcal{F})$, then, by Definition 7.1.1 and Definition 7.1.3, $g_2 = G(\beta)$ is the analytic continuation of $g_1 = G(\alpha)$ along some path γ from z_1 to z_2, and the lifted path Γ on $S(\mathcal{F})$ defined by (7.6) connects p_1 and p_2.

7.3 Typical Examples

Example 7.3.1 (The Riemann Surface of the Complex Logarithm). As a global analytic function, the *complex logarithm* $\mathcal{L}$ consists of all function elements (f, D), where D is a disk contained in $\mathbb{C} \setminus \{0\}$ and

$$f(z) = \operatorname{Log}|z| + \mathrm{i} \arg z,$$

with a continuous branch of the argument in D. The Riemann surface $S(\mathcal{L})$ consists of all points $p = (z, g)$, where z is a point in $\mathbb{C} \setminus \{0\}$ and g is a germ in $\mathcal{L}_z^*$. The value of any such germ can be expressed as

$$g(z) = \operatorname{Log} z + 2k\,\pi\,\mathrm{i} \tag{7.7}$$

with some integer k. When we denote points on $S(\mathcal{F})$ as $p = (z, g)$, the germ g only serves as an identifier for the different points p above z. Hence we can distinguish these points by labelling them by the integer k as well, so that $p = [z, k]$ is an alternative way of desribing the points on $S(\mathcal{L})$. In this notation, the lift of the logarithm to its Riemann surface is the function

$$\widetilde{\mathcal{L}} : S(\mathcal{L}) \to \mathbb{C},\ \ [z, k] \mapsto \operatorname{Log} z + 2k\pi\,\mathrm{i}.$$

We point out that this relabelling became possible by working with germs instead of function elements. While the germs at z are in a one-to-one correspondence with the integers k via (7.7), this does not hold for the functions f in the function elements (f, D) of $\mathcal{L}$ when the disk D intersects the negative real axis.

All points $p = [z, k]$ with fixed k form the kth sheet S_k of the Riemann surface $S(\mathcal{L})$. Every sheet S_k is a copy of the punctured complex plane, slit along the negative real axis $\mathbb{R}_-$. Note that $\mathbb{R}_-$ is not removed but attached to the ‘upper rim’ of the slit, which denotes the edge of the cut adjacent to the upper half plane.

A model of the Riemann surface of the logarithm is build by connecting the upper rim of the slit in S_k with the lower rim of the slit (adjacent to the lower half plane) in S_{k+1} as illustrated in Figure 7.6. Here the sheet on top is S_{-1} and S_0 is in the middle.

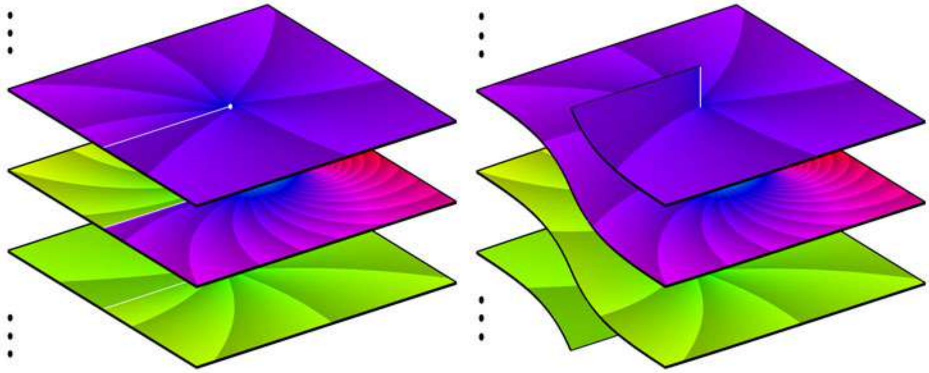


Figure 7.6: Building a model of the Riemann surface of the logarithm

Once we know the principles of construction, it is often more informative not to depict a function on a three-dimensional model of its Riemann surface, but on the projections of its sheets to the plane. Usually the phase portraits allow one to identify the branch cuts and the way in which their rims must be connected to form the Riemann surface. Figure 7.7 demonstrates this for three sheets of the logarithm.

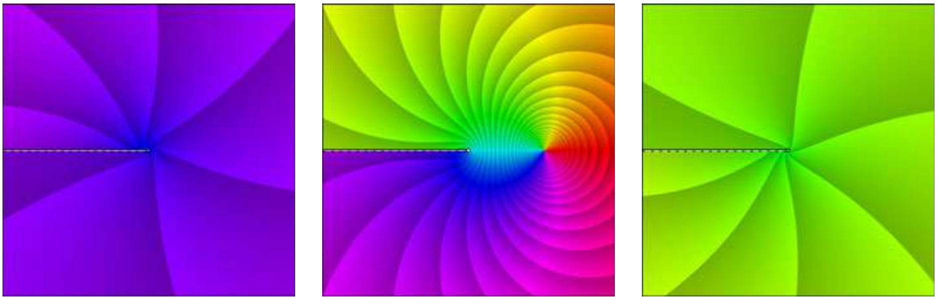


Figure 7.7: The logarithm on the sheets S_{-1}, S_0 and S_1 of its Riemann surface

The construction of a more involved Riemann surface is discussed at length in Section 7.6.

Example 7.3.2 (The Power Functions). The *power function* z^a with complex exponent $a \in \mathbb{C}$ is defined by

$$z^a := \mathrm{e}^{a \log z}.$$

As a global analytic function $\mathcal{Z}^a$ consists of all function elements $\big(\mathrm{e}^{a \log z}, D\big)$, where $(\log z, D)$ is a function element of the global logarithm. The structure of its Riemann surface depends heavily on the exponent a.

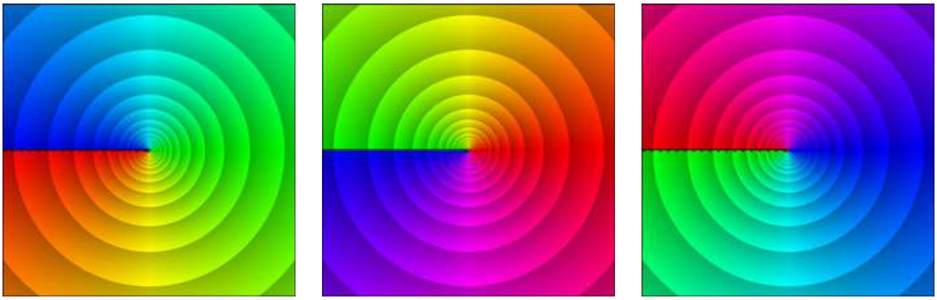


Figure 7.8: The function $f(z) = z^{2/3}$ on the three sheets of its Riemann surface

If a is a (real) *rational number*, $a = m/n$ with coprime integers m and n, then, due to the periodicity of the exponential function, $\mathscr{Z}^a$ has exactly n different germs at every point $z \in \mathbb{C} \setminus \{0\}$. Its Riemann surface can be build from n sheets $S_1, \ldots, S_n$ which are copies of $\mathbb{C} \setminus \{0\}$, with a branch cut along the negative real axis. The upper edge of S_k is connected with the lower edge of S_{k+1} for $k = 1, \ldots, n-1$ and the upper edge of S_n is joined with the lower edge of S_1.

If a is a (real) *irrational number*, then $\mathscr{Z}^a$ has infinitely many different germs at every point $z \in \mathbb{C} \setminus \{0\}$, and the structure of its Riemann surface is the same as for the logarithm.

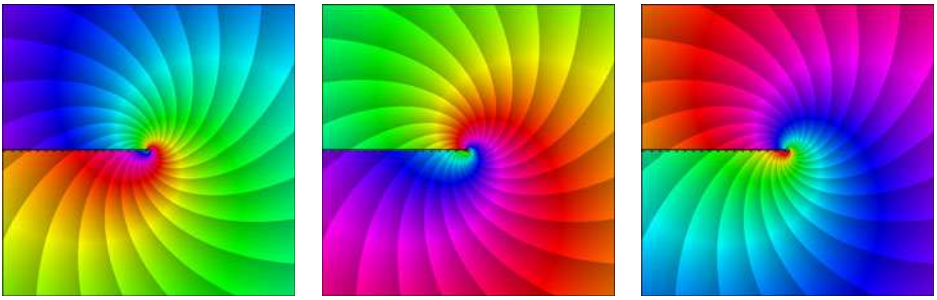


Figure 7.9: The function $f(z) = z^{2/3+\mathrm{i}}$ on three sheets of its Riemann surface

If a is not real, $a = \alpha + \mathrm{i}\beta$ with $\beta \neq 0$, then $\mathscr{Z}^a$ always has a countable number of germs at $z = r\mathrm{e}^{\mathrm{i}\varphi}$, which can be distinguished by their values at z,

$$g_k(z) = \mathrm{e}^{\alpha \operatorname{Log} r - \beta(\varphi + 2k\pi)}\, \mathrm{e}^{\mathrm{i}(\beta \operatorname{Log} r + \alpha(\varphi + 2k\pi))}, \quad k \in \mathbb{Z}.$$

The absolute values of $g_k(z)$ are all different and form a geometric progression with quotient $\mathrm{e}^{-2\pi\beta}$. Their arguments form an arithmetic progression with difference $\mathrm{e}^{2\pi\alpha}$. If α is rational, the phase of $g_k(z)$ is periodic in k. So, in that case, the phase portraits give the misleading impression that the Riemann surface consists only of a finite number of sheets. In particular, if a is purely imaginary, the global function z^a has the same phase portrait on all sheets of its Riemann surface. This is illustrated in Figure 7.10 for the function $f(z) = z^{\mathrm{i}}$ in the square $|\mathrm{Re}\, z| < 2$, $|\mathrm{Im}\, z| < 2$. It is a remarkable fact that all values of i^{i} are real.

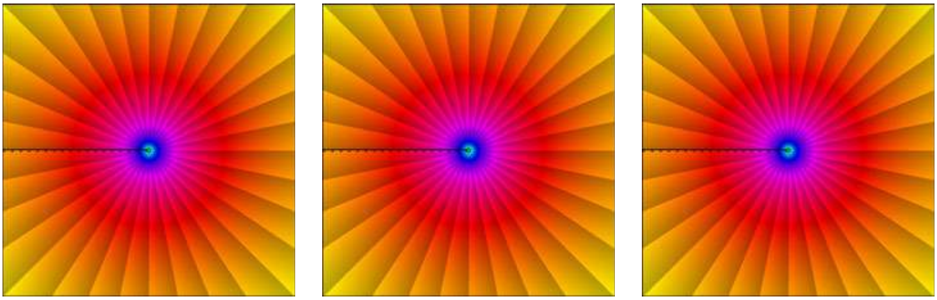


Figure 7.10: All sheets of $f(z) = z^{\mathrm{i}}$ have the same phase portrait

We mention that the unwanted coincidence of the phase of two functions f and g disappears when one replaces f and g by $f - c$ and $g - c$, respectively, where c is a non-zero constant c. To get a good contrast in some subdomain, one can choose c as the value of one of the functions f or g at a point in the domain of interest. In particular, barely visible branch cuts show up more clearly, as is demonstrated in Figure 7.11. Here the constant c is chosen as the value at $z = 1$ of the function depicted in the middle.

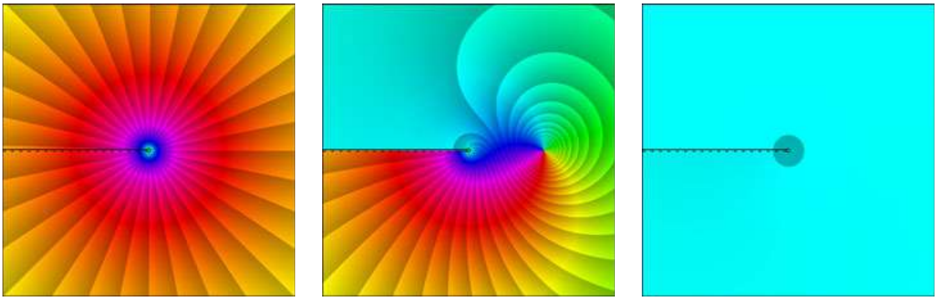


Figure 7.11: Phase portraits on three sheets of the modified function $f(z) - 1$

Since the modulus of z^{i} in the picture on the left is rather large, its phase is only slightly perturbed by subtracting -1. In the window on the right $|z^{\mathrm{i}}|$ is small, so that the constant -1 dominates.

Riemann Surfaces of Inverse Functions. A widespread source of Riemann surfaces is the construction of inverse functions. If an analytic function f is not injective its inverse is multiple-valued, and the *global inverse function* $\mathcal{F}^{-1}$ lives on a Riemann surface $S(\mathcal{F}^{-1})$. Usually $S(\mathcal{F}^{-1})$ can be directly constructed from sheets S_k corresponding to invertible restrictions of f to appropriately chosen subdomains. We demonstrate this in the next example.

Example 7.3.3 (The Arcsine Function). The complex sine function maps the strip $\{z \in \mathbb{C} : -\pi/2 < \operatorname{Re} z < \pi/2\}$ bijectively onto the plane slit along the segments $(-\infty, -1]$ and $[1, +\infty)$ (see Example 6.2.4). After complementing the strip by the segments

$$\{z \in \mathbb{C} : \operatorname{Re} z = -\pi/2,\ \operatorname{Im} z > 0\}, \qquad \{z \in \mathbb{C} : \operatorname{Re} z = \pi/2,\ \operatorname{Im} z < 0\}$$

to obtain a set G_0, we get a bijective mapping of G_0 onto $\mathbb{C} \setminus \{-1, 1\}$ as illustrated in the two pictures in the middle of Figure 7.12. The inverse of this mapping is the principal branch Arc sin of the *arcsine function.*

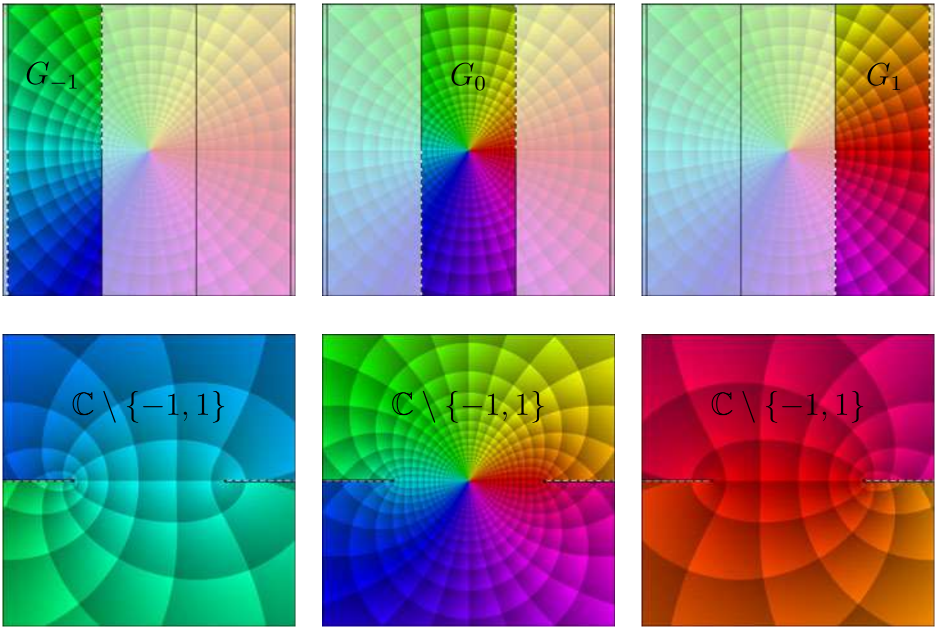

G_{-1}
G_0
G_1
$\mathbb{C} \setminus \{-1, 1\}$
$\mathbb{C} \setminus \{-1, 1\}$
$\mathbb{C} \setminus \{-1, 1\}$

Figure 7.12: Mapping properties of the sine and the arcsine on three sheets

Due to the identity $\sin(z + \pi) = -\sin z$, the sine function maps every translated strip $G_k := G + k\pi$ with $k \in \mathbb{Z}$ bijectively onto $\mathbb{C} \setminus \{-1, 1\}$. The corresponding inverse functions are

$$\operatorname{Arc\,sin}_k z = \begin{cases} \operatorname{Arc\,sin} z + k\pi & \text{if } k \text{ is even} \\ -\operatorname{Arc\,sin} z + k\pi & \text{if } k \text{ is odd.} \end{cases} \tag{7.8}$$

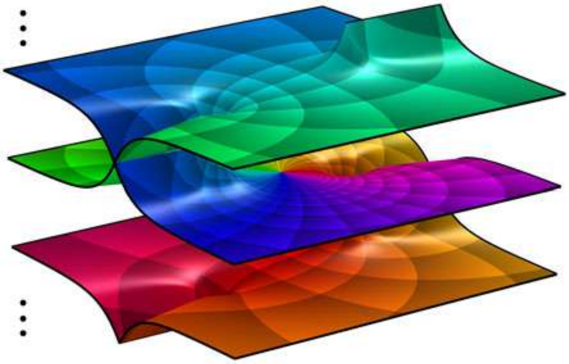


Figure 7.13: Riemann surface of the arcsine

The Riemann surface of the arcsine function has a countable number of sheets S_k each of which is a facsimile of $\mathbb{C} \setminus \{-1, 1\}$ slit along $(-\infty, -1)$ and $(1, +\infty)$. Both rims of the cut $(-\infty, -1)$ on S_{2k} are cross-connected with the edges of $(-\infty, -1)$ of S_{2k-1}, while the edges of $(1, +\infty)$ on S_{2k} are cross-connected with the edges of $(1, +\infty)$ of S_{2k+1}. Figure 7.13 shows three sheets of the resulting Riemann surface of the arcsine function.

7.4 Analytic Functions and Branch Points

Analytic Functions on a Riemann Surface. Once the basics are established, many concepts can be translated from the complex plane to Riemann surfaces. This book is not the place to pursue this further, we only exemplify how the notion of analytic functions extends to Riemann surfaces.

Let $S(\mathcal{F})$ be the Riemann surface of a global analytic function $\mathcal{F}$. Recall that a *chart* $\mathcal{P}_0$ of $S(\mathcal{F})$ at p_0 is a restriction of the canonical projection $\mathcal{P}$ to a disk on $S(\mathcal{F})$ centered at p_0.

Definition 7.4.1. A complex-valued function f defined in some neighborhood of a point p_0 on a Riemann surface $S(\mathcal{F})$ is said to be *analytic at* p_0, if for one (and then for all) charts $\mathcal{P}_0$ at p_0 the composition $f \circ \mathcal{P}_0^{-1}$ of f with the lifting map $\mathcal{P}_0^{-1}$ is analytic at $z_0 := \mathcal{P}(p_0)$. If f is defined and analytic at every point of $S(\mathcal{F})$ it is called *analytic on* $S(\mathcal{F})$.

Note that $f \circ \mathcal{P}_0^{-1}$ need only be defined in a neighborhood of z_0. The canonical projection $\mathcal{P} : (z, g) \mapsto z$ is analytic on $S(\mathcal{F})$ because $\big(\mathcal{P} \circ \mathcal{P}_0^{-1}\big)(z) = z$.

The next theorem confirms that the aim of converting a multiple-valued analytic function into a single-valued analytic function is achieved by lifting a global analytic function to its Riemann surface.

Theorem 7.4.2. *The lift* $\widetilde{\mathcal{F}}$ *of any global analytic function* $\mathcal{F}$ *is analytic on its Riemann surface* $S(\mathcal{F})$.

Proof. If $\mathcal{P}_0 : \widetilde{D}_0 \to D_0$ is a chart at $p_0 = (z_0, g_0)$, the disk $\widetilde{D}_0$ is associated with a function element (f_0, D_0) in $\mathcal{F}$ and we have $\widetilde{\mathcal{F}} \circ \mathcal{P}_0^{-1} = f_0$ on D_0 (see (7.4)). $\square$

Isolated Singularities. Let us now review the construction of the Riemann surfaces of the square root and the logarithm in Examples 7 and 7.3.1. In both examples the origin plays an exceptional role, in that the function elements which generate the Riemann surface admit an unrestricted analytic continuation to $\mathbb{C} \setminus \{0\}$, but not to $\mathbb{C}$. Despite this common property, the local structure of the corresponding Riemann surfaces $S(\mathcal{F})$ and the behavior of the functions $\widetilde{\mathcal{F}}$ in a neighborhood of this "singular point" are rather different. In order to explore this further, we make the following definition.

Definition 7.4.3. A point $a \in \mathbb{C}$ is called an *isolated singularity* of the global analytic function $\mathcal{F}$ if there exist a punctured disk $\dot{D} := D \setminus \{a\}$ and a function element (f_0, D_0) in $\mathcal{F}$ with $D_0 \subset \dot{D}$ which admits an unrestricted analytic continuation in $\dot{D}$, but has no unrestricted analytic continuation in D.

In what follows we assume that $\dot{D}$ and (f_0, D_0) satisfy the conditions in Definition 7.4.3, and analyze the structure of the Riemann surface $S(\mathcal{F})$ above the punctured disk $\dot{D}$, viz. the set $\mathcal{P}^{-1}(\dot{D})$. First of all we observe that $\mathcal{P}^{-1}(\dot{D})$ need not be connected but may consist of several components. Two points $p_1 = (z_1, g_1)$ and $p_2 = (z_2, g_2)$ belong to the same component if and only if the germ g_2 is the analytic continuation of g_1 along a path in $\dot{D}$. Since we are only interested in the *local structure* of $S(\mathcal{F})$, the components of $\mathcal{P}^{-1}(\dot{D})$ can be studied separately. We therefore consider the Riemann surface $S(\mathcal{F}_0)$ of the global analytic function $\mathcal{F}_0$ generated by a function element (f_0, D_0) through analytic continuation along paths in $\dot{D}$.

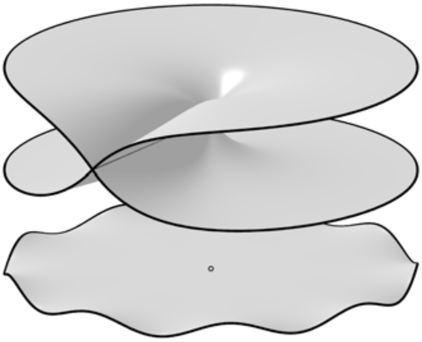


Figure 7.14: Components of $\mathcal{P}^{-1}(\dot{D})$

The next theorem describes the simplest situation. We state it in the language of germs and in a slightly more general context than needed. Recall that the analytic continuation $g(\gamma)$ of a germ g along a closed path γ is said to be trivial if $g(\gamma) = g$.

Theorem 7.4.4. *Let G be a simply connected domain, assume that the germ f_0^* has an unrestricted analytic continuation in the punctured domain $\dot{G} := G \setminus \{a\}$, and let $\mathcal{F}_0$ be the global analytic function generated by f_0^* in $\dot{G}$. If there exist a closed path γ in $\dot{G}$ with winding number one about the point a, and a germ in $\mathcal{F}_0^*$ with trivial analytic continuation along γ, then all germs in $\mathcal{F}_0^*$ are induced by an analytic function $f : \dot{G} \to \mathbb{C}$.*

Proof. The result is an immediate corollary of Theorem 3.6.10 (iii). ☐

Rephrased as a statement about Riemann surfaces, Theorem 7.4.4 just tells us that $S(\mathcal{F}_0)$ consists of a single sheet above $\dot{G}$.

Branch Points. We now explore the more interesting situations in which the assumptions of Theorem 7.4.4 are not satisfied.

Definition 7.4.5. Let $\mathcal{F}_0$ be a global analytic function in a disk D and assume that $a \in D$ is the only isolated singularity of $\mathcal{F}_0$ in D. Then a is said to be a *branch point* of $\mathcal{F}_0$, if one (and then any) germ in $\mathcal{F}_0^*$ has a non-trivial analytic continuation along some closed path in $\dot{D} := D \setminus \{a\}$.

Let $\mathcal{F}_0$ be generated by a germ f_0^* at $z_0 \in \dot{D}$. Since the analytic continuation of f_0^* along a closed path γ in $\dot{D}$ depends only on the winding number k of γ about a (see Lemma 2.7.22 and Theorem 3.6.8), we can label the germs in $\mathcal{F}_0^*$ at z_0 so that f_k^* is the analytic continuation of f_0^* along a path with winding number k about a. Then there are two possibilities:

(i) The germs f_k^* and f_m^* are different whenever $k \neq m$.

(ii) The germs f_k^* and f_m^* are equal for some $m > k$.

In the first case we say that a is a *branch point of infinite order*. Since this case happens for the isolated singularity of the logarithm, we also speak of a *logarithmic branch point*.

In the second case, analytic continuation of f_m^* and f_k^* along a path with winding number $-k$ shows that (ii) implies $f_{m-k}^* = f_0^*$. Consequently, there is a smallest number $n \geq 2$ (because $n = 1$ has been excluded) such that $f_n^* = f_0^*$, and then the mapping $k \mapsto f_k^*$ is periodic with primitive period n, so that $\mathcal{F}_0^*$ contains exactly n different germs $f_0^*, \dots, f_{n-1}^*$ at z_0. In this case a is said to be a *branch point of order n*.

It can easily be seen that the choice of f_0^* in $\mathcal{F}_0$ does not influence which of the cases (i) or (ii) appears and what the value of n is.

Summarizing the results we get the following theorem on the classification of isolated singularities of global analytic functions.

Theorem 7.4.6. *Let G be a simply connected domain and assume that a germ f_0^* in the punctured domain $\dot{G} := G \setminus \{a\}$ has an unrestricted analytic continuation in $\dot{G}$. Then the global analytic function $\mathcal{F}_0$ which is generated in $\dot{G}$ by f_0^* has exactly one of the following properties:*

(i) *$\mathcal{F}_0^*$ is induced by an analytic function $f : \dot{G} \to \mathbb{C}$.*

(ii) *$\mathcal{F}_0^*$ has a branch point of finite order n at a.*

(iii) *$\mathcal{F}_0^*$ has a logarithmic branch point at a.*

Example 7.4.1 (The Root Functions). The *root function* $\sqrt[n]{z}$ has an isolated singularity at the origin which is a branch point of order n. All points $p = (z, g_k)$

on its Riemann surface $S(\sqrt[n]{z}\,)$ have the form $p=(z,g_k)$ with $k=1,\dots,n$, where $z\in\mathbb{C}\setminus\{0\}$ and g_k is the germ of $\sqrt[n]{z}$ specified by its value at z,

$$g_k(z)=\exp\left(\frac{1}{n}\left(\operatorname{Log} z+2(k-1)\pi\,\mathrm{i}\right)\right). \tag{7.9}$$

Like for the logarithm (see Example 7.3.1), this yields the alternative representation $p=[z,k]$ for points p on the kth sheet of $S(\sqrt[n]{z}\,)$.

Local Normal Forms. The Riemann surfaces of the root functions and the logarithm are local models for the Riemann surfaces of general global analytic functions in a neighborhood of branch points. For example, any global analytic function $\mathcal{F}_0$, which is generated in the punctured disk $\dot{D}=D\setminus\{a\}$ by a germ f_0^* at z_0, and which has a branch point of order n at a, can be lifted to an analytic function $\widetilde{\mathcal{F}}$ on the Riemann surface of $\sqrt[n]{z-a}$ above $\dot{D}$. This construction is described next.

1. We define the function φ in the punctured disk $\dot{G}:=\{w\in\mathbb{C}:\ a+w^n\in\dot{D}\}$ by $\varphi(w):=a+w^n$ and choose $w_0\in\dot{G}$ such that $\varphi(w_0)=z_0$. Then the composition $f_0^*\circ\varphi$ of φ with the germ f_0^* is a germ g_0^* at w_0: if (f_0,D_0) represents f_0^*, the function $g_0:=f_0\circ\varphi$ is defined and analytic in a neighborhood of w_0, and g_0^* is the germ at w_0 induced by g_0.

2. We show that g_0^* has an unrestricted analytic continuation in $\dot{G}$. Indeed, if γ is a path in $\dot{G}$, then $\varphi\circ\gamma$ is a path in $\dot{D}$. If f^* is the analytic continuation of f_0^* along $\varphi\circ\gamma$, then the analytic continuation g^* of g_0^* along γ is the composition $g^*=f^*\circ\varphi$.

3. If γ is a closed path in $\dot{G}$ with winding number 1 about the origin, then the corresponding path $\varphi\circ\gamma$ in $\dot{D}$ has winding number n about a. Since a is a branch point of order n of $\mathcal{F}_0$, the analytic continuation of f_0^* along $\varphi\circ\gamma$ is trivial, and hence the same holds for the analytic continuation of g_0^* along γ. So g_0^* satisfies the assumptions of Theorem 7.4.4, and thus it generates an analytic function g in the domain $\dot{G}$.

4. Finally, we denote by $\mathcal{R}$ the root function $\sqrt[n]{z-a}$ on its Riemann surface S above $\dot{D}$, and define $\widetilde{\mathcal{F}}_0$ on S by

$$\widetilde{\mathcal{F}}_0:=g\circ\mathcal{R}. \tag{7.10}$$

More explicitly, writing the points on the kth sheet S_k of S as $[z,k]$ with $z\in\dot{D}$ for $k=1,\dots,n$ (see Example 7.4.1), we get the representation

$$\widetilde{\mathcal{F}}_0:S\to\mathbb{C},\ [z,k]\mapsto g\big(\varrho_k(z)\big),$$

where ϱ_k denotes the kth branch of the root function $\sqrt[n]{z-a}$ in $\dot{D}$,

$$\varrho_k(z)=\exp\left(\frac{1}{n}\left(\operatorname{Log}(z-a)+2(k-1)\pi\mathrm{i}\right)\right). \tag{7.11}$$

So $f_k := g \circ \varrho_k$ is the restriction of $\widetilde{F}_0$ to the sheet S_k of S.

The function $\widetilde{\mathcal{F}}_0$ is analytic on S in the sense of Definition 7.4.1 and lifts the global analytic function $\mathcal{F}_0$ to the Riemann surface S: every function element (f_1, D_1) in $\mathcal{F}_0$ has the form

$$\big(\widetilde{\mathcal{F}}_0 \circ \mathcal{P}_1^{-1}, D_1\big), \tag{7.12}$$

where $\mathcal{P}_1^{-1} : D_1 \to \widetilde{D}_1$ lifts the disk D_1 to some disk $\widetilde{D}_1$ on S, and, conversely, any such function element belongs to $\mathcal{F}_0$.

Rephrased in compact form, we have shown that any function f with a branch point of order n at a can locally be represented in the *normal form*

$$f(z) = g\big(\sqrt[n]{z-a}\,\big), \tag{7.13}$$

where g is analytic in a punctured disk $\dot{G} = G \setminus \{0\}$. We shall call g the *analytic function associated with the branch point* a of f. Of course, formula (7.13) can also be interpreted in the framework of multiple-valued functions.

The construction is illustrated in Figure 7.15. The point of departure is the green disk in the lower left picture, which is the phase portrait of a function element (f_0, D_0) representing the germ f_0^*. It generates a global analytic function $\mathcal{F}_0$ with a branch point of second order at the origin. With the chessboard-like structure of the phase portrait we intended to simultaneously visualize the two branches of the corresponding multiple-valued function.

The pull-back of the germ f_0^* via $\varphi(w) = a + w^n$ (with $a = 0$ and $n = 2$) is a germ g_0^* at w_0. It is represented by the function element (g_0, G_0), displayed in the green disk in the lower right window (w-plane).

In order to construct the analytic continuation of g_0^* along the closed path γ (with winding number one), γ is transplanted to the path $\widetilde{\gamma} := \gamma \circ \varphi$ (with winding number two) in the z-plane. Since the branch point of $\mathcal{F}_0$ has order two, the analytic continuation of f_0^* along $\widetilde{\gamma}$ is trivial, which implies that g_0^* has a trivial analytic continuation along γ, too. Consequently (g_0, G_0) generates an analytic function g in $\dot{G}$. The phase portrait of the function g is depicted in the lower right window; it is generated by pulling back the phase portrait in the upper right window.

Finally, the function g is transplanted to the Riemann surface S of the square root. The upper left picture visualizes the surface S, carrying the phase portrait of $\widetilde{\mathcal{F}}_0 := g \circ \mathcal{R}$. This phase portrait is the pull back of the phase portrait of g in the lower right window via the square root $\mathcal{R}$ on S. The black line on S is the lifting of the path $\widetilde{\gamma}$ via the canonical projection $\mathcal{P}$.

The function elements in $\mathcal{F}_0$ are exactly the projections of the restrictions of $\widetilde{\mathcal{F}}_0$ to the disks on S. This is demonstrated in the two pictures on the left for three elements, among them the generating element (f_0, D_0) of $\mathcal{F}_0$.

The phase portraits of the branches $f_1 := g \circ \varrho_1$ and $f_2 := g \circ \varrho_2$ on the two sheets S_1 and S_2 of S are depicted in Figure 7.16. The relations between the phase portraits of the associated function g and the branches f_k are visible more clearly in Figure 7.17 for a branch point of order three.

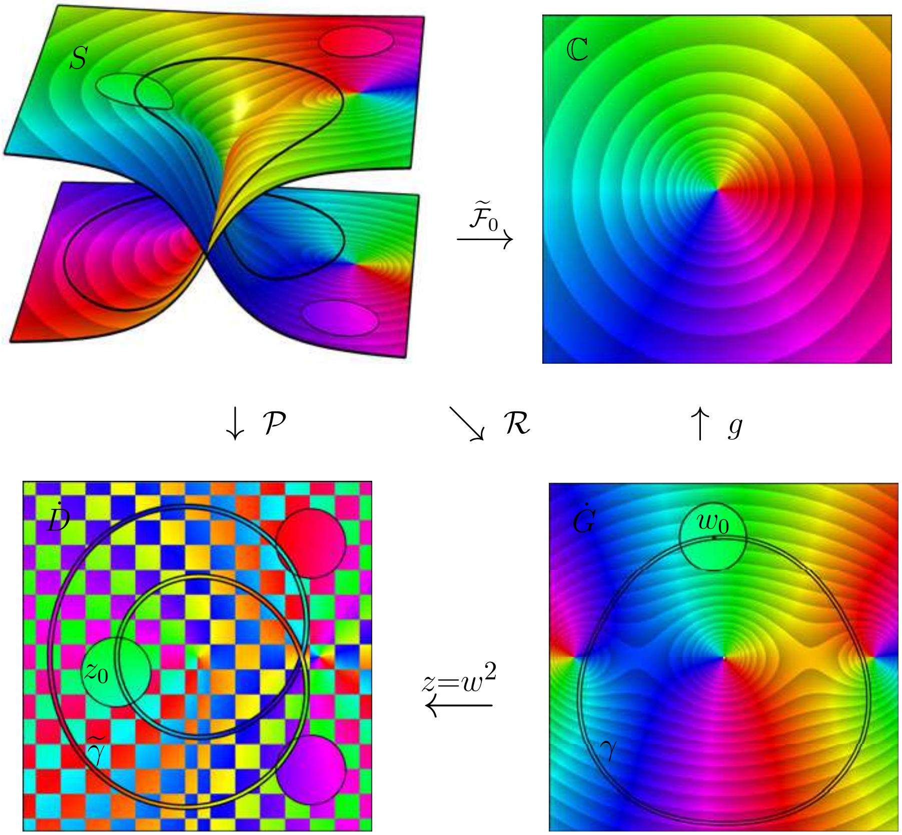


Figure 7.15: Lifting a multiple-valued function to a Riemann surface

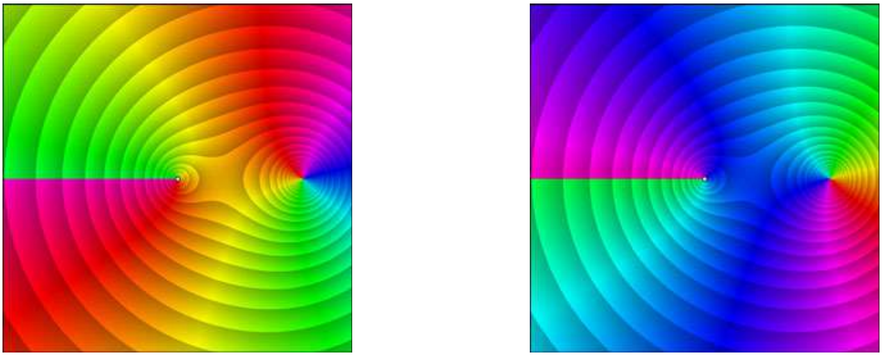


Figure 7.16: The functions $g \circ \varrho_1$ and $g \circ \varrho_2$ on the sheets S_1 and S_2 of S

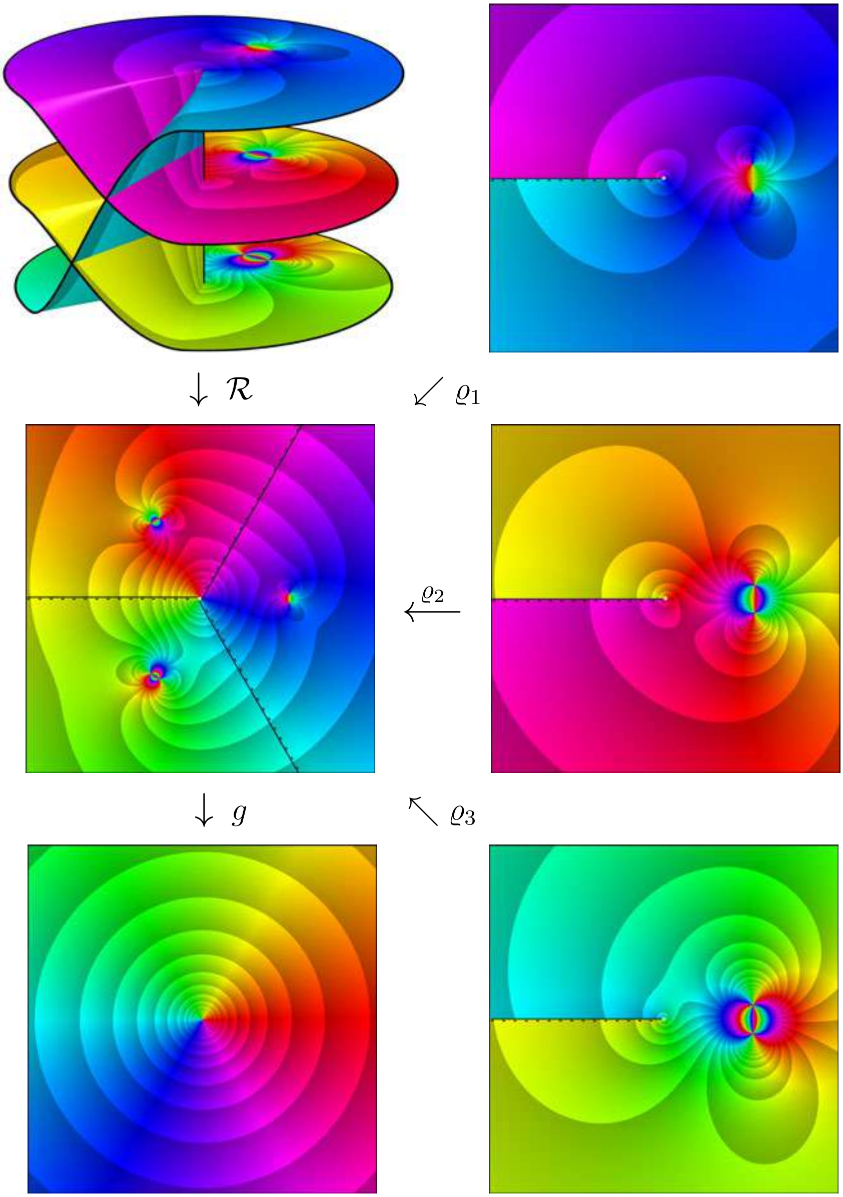


Figure 7.17: A branch point of order three on a Riemann surface and its sheets

According to Theorem 4.4.1, the associated analytic function g in the punctured disk $\dot{G}$ can be represented by a *Laurent series*,

$$g(z) = \sum_{k=-\infty}^{\infty} c_k \, z^k, \qquad z \in \dot{G}. \tag{7.14}$$

In conjunction with (7.13) this yields a representation of f by a *Puiseux series* involving *fractional powers*,

$$f(z) = \sum_{k=-\infty}^{\infty} c_k \, (z-a)^{k/n}, \qquad z \in \dot{D}. \tag{7.15}$$

More precisely, if (f, D_0) is a function element in $\mathcal{F}_0$, then f can be represented by (7.15) with an appropriately chosen analytic branch of $(z-a)^{1/n}$ in D_0.

Regular Branch Points. If the singularity of the associated analytic function g at the center of the punctured disk $\dot{G}$ is removable, we speak of a *regular branch point*. For example, the root functions $\sqrt[n]{z}$ have regular branch points at the origin, since then $g(z) = z$.

Regular branch points of $\mathcal{F}_0$ can be incorporated in a natural way into the Riemann surface $S(\mathcal{F}_0)$. Informally, the *extended Riemann surface* $\widehat{S}(\mathcal{F}_0)$ is built by complementing $S(\mathcal{F}_0)$ with a single point p_0 above a, at which all sheets are glued together.

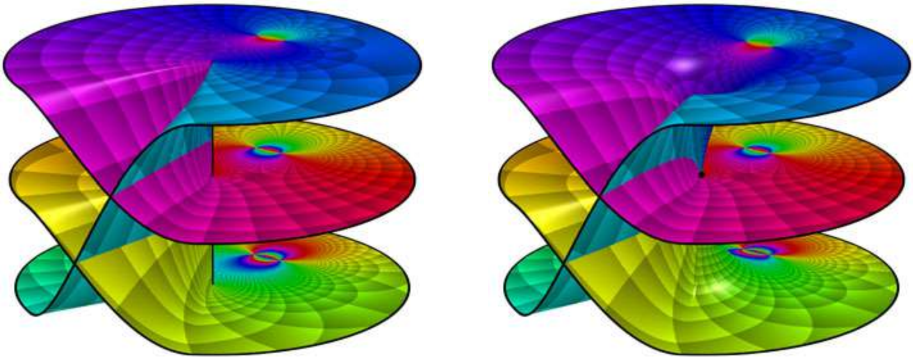


Figure 7.18: Compactification of a Riemann surface at a regular branch point

Since g has a removable singularity at 0, the function $\widetilde{\mathcal{F}}_0$ has a limit at p_0, so that it can be extended to a continuous function $\widehat{\mathcal{F}}_0$ on $\widehat{S}(\mathcal{F}_0)$. In fact this function is not only continuous but even *analytic* at p_0, at least when we use a more general notion of Riemann surfaces and analytic functions. This brings us to the final step of abstraction.

7.5 Abstract Riemann Surfaces

Leaving aside all technical details, only two essential ingredients of a Riemann surface remain: a *topological space* on which the surface is modelled *globally*, and some *complex structure*, which makes this space look locally like the complex plane.

The first requirement is *a-priori* satisfied when we build the Riemann surface on a Hausdorff topological space $\mathbb{S}$. The complex (or conformal) structure on $\mathbb{S}$ is then generated by local charts. A (local) *chart* on $\mathbb{S}$ is a homeomorphic mapping φ from an open subset U of $\mathbb{S}$ onto an open set in the complex plane. To have access to both, the homeomorphism φ and its domain U, we denote charts by (U, φ). An *atlas* for $\mathbb{S}$ is a family of charts $(U_\alpha, \varphi_\alpha)$ (with α from some unspecified index set) such that $\mathbb{S}$ is covered by the domain sets U_α, i.e., any point p on $\mathbb{S}$ is contained in some U_α.

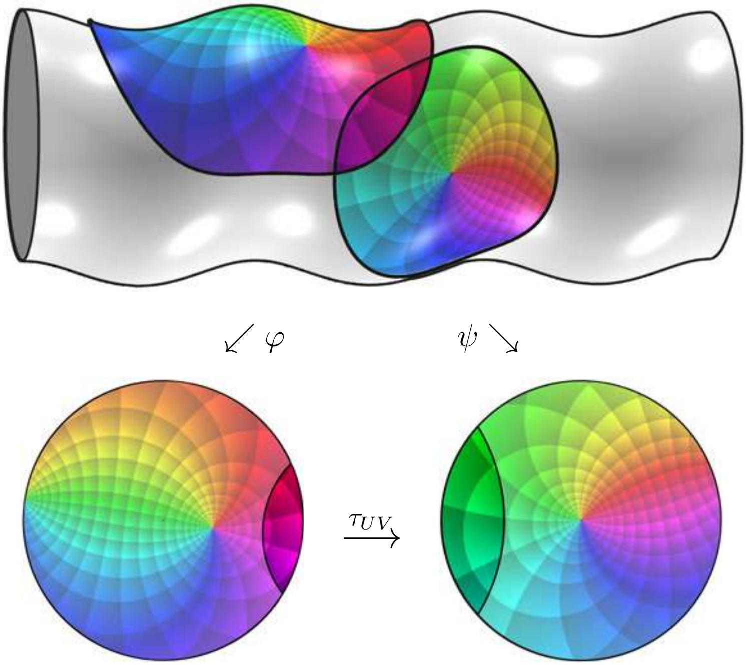


Figure 7.19: Two charts on a Riemann surface and their transition map

Every chart $(U_\alpha, \varphi_\alpha)$ pulls back the structure of the complex plane to its domain U_α. When we wish to glue all these local structures to some global entity, certain compatibility conditions must be satisfied in the overlapping regions. In order to obtain these conditions we pick two charts (U, φ) and (V, ψ) and consider the *transition mapping*

$$\tau_{UV} : \varphi(U \cap V) \to \mathbb{C}, \; z \mapsto \psi\big(\varphi^{-1}(z)\big), \tag{7.16}$$

which is defined on the open (possibly empty) subset $\varphi(U \cap V)$ of the complex plane. Two charts (U, φ) and (V, ψ) are said to be *compatible* if either $U \cap V = \emptyset$ or both transition mappings τ_{UV} and τ_{VU} are analytic in their domains. If all maps of an atlas are compatible, we speak of a *conformal atlas*.

Definition 7.5.1. An *abstract Riemann surface* $\mathbb{S}$ (or a *one-dimensional complex manifold*) is a Hausdorff space endowed with a conformal atlas.

Example 7.5.1. Any open subset D of the *complex plane*, endowed with the Euclidean topology and the atlas consisting of the single chart (D, φ) with $\varphi(z) := z$ is an abstract Riemann surface.

Sometimes it is convenient to work with more specific charts (U, φ). For example, one may suppose that the image domains $\varphi(U)$ are disks. In the next example all charts are mappings onto the unit disk.

Example 7.5.2. Another conformal atlas (U_a, φ_a) on a (proper) open subset D of $\mathbb{C}$ is given by

$$U_a := \{z \in \mathbb{C} : |z - a| < r_a\}, \quad \varphi_a(z) := (z - a)/r_a, \qquad a \in D, \tag{7.17}$$

where U_a is the largest disk centered at a which is contained in D. All transition maps are linear functions and hence they are analytic.

Equivalent Atlases. The two examples above show that one and the same underlying topological space can be equipped with different conformal atlases. Though the two resulting Riemann surfaces are formally different, these differences are not essential. The following definition makes this precise and eliminates the dependence of a Riemann surface on the special choice of an atlas.

Definition 7.5.2. Two conformal atlases on $\mathbb{S}$ are said to be *equivalent* if their union is a conformal atlas.

It is clear that equivalence of atlases is an equivalence relation, and a class of equivalent atlases is called a *conformal structure*. As long as atlases are equivalent, it does not really matter which one we choose, they all induce the same conformal structure. Therefore we shall often identify Riemann surfaces when their atlases are equivalent.[2] The following two examples illustrate an even more general concept of equivalence which will be introduced afterwards.

Example 7.5.3 (The Extended Complex Plane). The *extended complex plane* is the set $\mathbb{C} \cup \{\infty\}$, endowed with the topology induced by stereographic projection from the topology of the unit sphere in $\mathbb{R}^3$. It has a conformal atlas with two charts

$$\begin{aligned} U_0 &:= \{z \in \mathbb{C} : |z| < 2\}, & \varphi_0(z) &:= z, \\ U_\infty &:= \{z \in \mathbb{C} \cup \infty : |z| > 1/2\}, & \varphi_\infty(z) &:= 1/z. \end{aligned}$$

---

[2] An alternative definition *introduces* Riemann surfaces as Hausdorff spaces endowed with a conformal structure.

Both transition maps are analytic in the ring $R := \{z \in \mathbb{C} :\ 1/2 < |z| < 2\}$,

$$(\varphi_0 \circ \varphi_\infty^{-1})(z) = (\varphi_\infty \circ \varphi_0^{-1})(z) = 1/z.$$

Example 7.5.4 (The Unit Sphere). The *unit sphere* $\mathbb{S}$ in the Euclidean space $\mathbb{R}^3$ has a conformal atlas consisting of two charts

$$U_+ := \{(X,Y,Z) \in \mathbb{S} :\ Z < 1/2\}, \qquad \varphi_+(X,Y,Z) := \frac{X + \mathrm{i}\,Y}{1-Z},$$
$$U_- := \{(X,Y,Z) \in \mathbb{S} :\ Z > -1/2\}, \qquad \varphi_-(X,Y,Z) := \frac{X - \mathrm{i}\,Y}{1+Z}.$$

The mappings φ_+ and φ_- are the stereographic projections from the north pole $(0,0,1)$ and the south pole $(0,0,-1)$ onto the upper and the lower side of the equatorial plane, respectively. The minus sign in the numerator of φ_- appears since we address points on both sides of that plane using the same coordinates $z = x + \mathrm{i}y$. Both transition maps are analytic in the ring $R := \{z \in \mathbb{C} :\ 1/\sqrt{3} < |z| < \sqrt{3}\}$: taking into account that $X^2 + Y^2 + Z^2 = 1$ on $\mathbb{S}$, we have

$$\frac{1}{\varphi_+(X,Y,Z)} = \frac{1-Z}{X+\mathrm{i}\,Y} = \frac{(1-Z)(X-\mathrm{i}\,Y)}{X^2+Y^2} = \frac{X-\mathrm{i}\,Y}{1+Z} = \varphi_-(X,Y,Z),$$

so that

$$(\varphi_+ \circ \varphi_-^{-1})(z) = (\varphi_- \circ \varphi_+^{-1})(z) = 1/z, \quad z \in R.$$

When we identify the extended complex plane $\mathbb{C} \cup \{\infty\}$ with the sphere $\mathbb{S}$ by stereographic projection, the two atlases in Example 7.5.3 and Example 7.5.4 are equivalent. So both can be considered as models of one and the same Riemann surface, the Riemann sphere $\widehat{\mathbb{C}}$. This observation leads to a more general notion of equivalence which we shall discuss later.

Example 7.5.5 (The Riemann Surfaces $S(\mathcal{F})$). The Riemann surface $S(\mathcal{F})$ of a global analytic function $\mathcal{F}$ in the sense of Definition 7.1.1 is an abstract Riemann surface. We have already seen in Section 7.2 that the set $S(\mathcal{F})$ defined in (7.1) carries the topology of a Hausdorff space. A conformal atlas is given by the charts $\big(\widetilde{D}_p, \varphi_p\big)$, labelled by the points p in $S(\mathcal{F})$, where $\widetilde{D}_p$ is a disk on $S(\mathcal{F})$ centered at p, and

$$\varphi_p : \widetilde{D}_p \to D_p \subset \mathbb{C},\ (z,g) \mapsto z$$

is the restriction of the canonical projection $\mathcal{P}$ to $\widetilde{D}_p$.

Analytic Mappings. The concept of analytic functions can easily be extended to mappings between abstract Riemann surfaces. The idea is to 'project' functions locally from the Riemann surface to open sets in the complex plane, using the charts of a conformal atlas. If $f : \mathbb{S} \to \widetilde{\mathbb{S}}$ is a continuous mapping between two Riemann surfaces, we choose charts (U,φ) in the atlas of $\mathbb{S}$, and (V,ψ) in the atlas of $\widetilde{\mathbb{S}}$ and consider the composition

$$f_{UV} := \psi \circ f \circ \varphi^{-1} :\ D_{UV} \to \mathbb{C}. \tag{7.18}$$

Since f is assumed continuous, the domain of definition $D_{UV} := \varphi\big(U \cap f^{-1}(V)\big)$ of f_{UV} is a (possibly empty) open set in the plane.

Definition 7.5.3. A continuous mapping $f : \mathbb{S} \to \widetilde{\mathbb{S}}$ between Riemann surfaces is said to be *analytic* if all functions (7.18) with $D_{UV} \neq \emptyset$ are analytic.

Note that this definition is compatible with Definition 3.5.1, where we included the point at infinity in the domain and the range of analytic functions. For example, according to Definition 3.5.1 a function $f : D \subset \widehat{\mathbb{C}} \to \mathbb{C}$ is analytic at $\infty \in D$ if $g : z \mapsto f(1/z)$ is analytic at the origin. The function g is just the composition $\psi \circ f \circ \varphi_\infty^{-1}$, where $(U_\infty, \varphi_\infty)$ with $\varphi_\infty(z) := 1/z$ is a chart of $\widehat{\mathbb{C}}$ at infinity (see Example 7.5.3), and $(\mathbb{C}, \psi)$ with $\psi(z) := z$ is a (global) chart on $\mathbb{C}$.

Example 7.5.6 (The Root Functions on the Riemann Sphere). In Example 7.4.1 we have constructed the Riemann surfaces of the root functions $\sqrt[n]{z}$ on $\mathbb{C}\setminus\{0\}$. Recall that the points p on such a Riemann surface S can be represented as $p = [z, k]$, where k refers to the number of the sheet of S on which p is located. So, considering the Riemann surface just as a *set*, we have

$$S = \{[z, k] : z \in \mathbb{C} \setminus \{0\},\ k = 1, \dots, n\}.$$

According to Example 7.5.5, a *conformal atlas* (which also induces the topology) on S is given by the charts (U_p, φ_p), labelled by the points $p = [z, k]$ in S, where U_p is the disk in S with center p and radius $|z|$, and

$$\varphi_p : U_p \to \mathbb{C},\ [z, k] \mapsto z. \tag{7.19}$$

However, there is a much simpler equivalent atlas, consisting of the single chart (S, σ), where $\sigma : S \to \mathbb{C}$ is the (lifted) root function $\sqrt[n]{z}$ on S. This chart is a homeomorphism between S and $\mathbb{C} \setminus \{0\}$, and its compatibility with the old atlas follows because $\sigma \circ \varphi_p^{-1}$ is an analytic branch of the nth root on $\varphi_p(U_p)$ (by Theorem 7.4.2 σ is analytic on S), and $\varphi_p \circ \sigma^{-1} : z \mapsto z^n$ is analytic on $\sigma(U_p)$.

In order to incorporate the branch point at the origin into S we first complement the *set* S by a point $p_0 = [0, 0]$. The left zero in this notation indicates that this point will lie above $z = 0$ on the extended Riemann surface. The right zero tells us that p_0 does not belong to any of the sheets $S_1, \dots, S_n$.

The *canonical projection* $\mathcal{P}$ on $\overline{S} := S \cup \{[0, 0]\}$ is defined consistently with the projection on S by

$$\mathcal{P} : \overline{S} \to \mathbb{C},\ [z, k] \mapsto z.$$

The *topology* on $\overline{S}$ is introduced by complementing the topology of $\overline{S}$ by a neighborhood basis of $[0, 0]$, consisting of the preimages $\mathcal{P}^{-1}(D)$ of all disks $D \subset \mathbb{C}$ centered at the origin. The intuitive meaning of this construction is that the point $[0, 0]$ is now attached to all sheets of S, so that they are linked at this point.

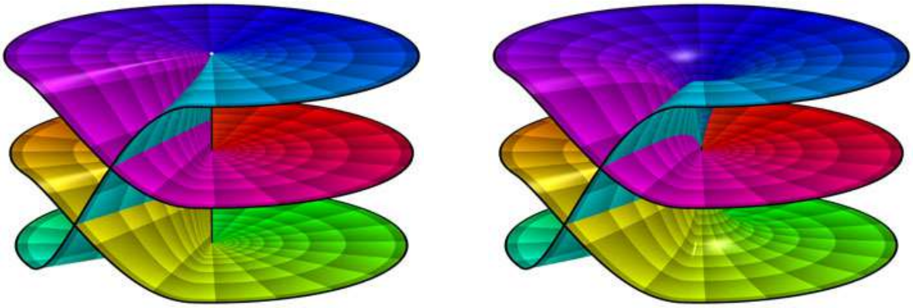


Figure 7.20: The Riemann surfaces S and $\overline{S}$ of the cubic root above $\mathbb{D}$

The extension of the atlas on $\mathcal{S}$ to an atlas on $\overline{S}$ is quite simple: we add a single chart which covers the complete surface $\overline{S}$, namely $(\overline{S}, \overline{\sigma})$, where $\overline{\sigma}$ is the root function on S, extended to $\overline{S}$ by $\sigma([0,0]) := 0$. It is clear that $\overline{\sigma}$ is a homeomorphism of $\overline{S}$ onto $\mathbb{C}$, and it is trivial that the two charts (S, σ) and $(\overline{S}, \overline{\sigma})$ are compatible. Once the chart $(\overline{S}, \overline{\sigma})$ is established, we can of course forget about the subordinate chart (S, σ).

In a similar manner, $\overline{S}$ can be complemented to a Riemann surface $\widehat{S}$ by adjoining a point $[\infty, \infty]$ which forms the fiber at $z = \infty$. Furthermore, the root function on $\overline{S}$ is continuously extended to a function $\widehat{\sigma} : \widehat{S} \to \widehat{\mathbb{C}}$ by setting $\widehat{\sigma}([\infty, \infty]) := \infty$. Note that $\widehat{\sigma}$ is a homeomorphism. A conformal atlas on $\widehat{S}$ is then given by the two charts

$$\big(\widehat{S} \setminus \{[\infty, \infty]\}, \widehat{\sigma}\big), \qquad \big(\widehat{S} \setminus \{[0,0]\}, 1/\widehat{\sigma}\big).$$

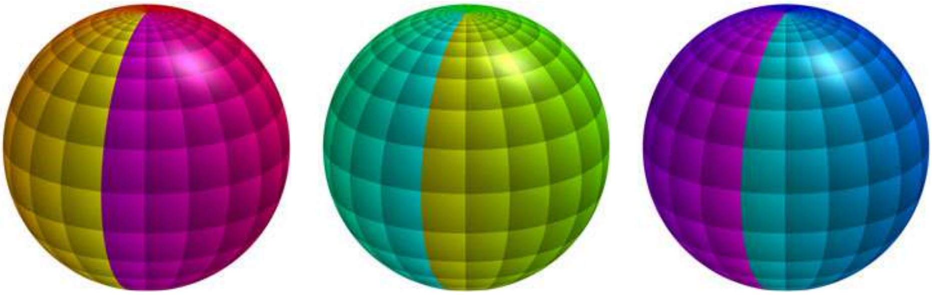


Figure 7.21: The cubic root on three sheets of $\widehat{S}$ covering $\widehat{C}$

The addition of the two points $[0,0]$ and $[\infty, \infty]$ turns the non-compact surface S into a compact Riemann surface $\widehat{S}$, which is a threefold covering of the Riemann sphere. Figure 7.21 shows the enhanced phase portraits of the cubic root on the three sheets of $\widehat{S}$ covering $\widehat{\mathbb{C}}$.

Local Normal Form at Regular Branch Points. As we have seen above, any global analytic function $\mathcal{F}$ in a punctured disk $\dot{D} := D \setminus \{a\}$ with a branch point of order n at a can be lifted to an analytic function $\widetilde{\mathcal{F}}$ on the Riemann surface $S := S(\sqrt[n]{z-a}\,)$ of the nth root above D.

Theorem 7.5.4. *If a is a* regular *branch point of order n of a global analytic function $\mathcal{F}$ in a punctured disc $\dot{D}$, then the lifted function $\widetilde{\mathcal{F}}$ has an analytic continuation $\widehat{\mathcal{F}}$ to the extended Riemann surface $\widehat{S}$ of $\sqrt[n]{z-a}$ above D.*

Proof. If $g : \dot{G} \to \mathbb{C}$ denotes the analytic function associated with the branch point a, we have $\widetilde{F} := g \circ \sigma$, with $\sigma(z) = \sqrt[n]{z-a}$ on S (see page 326). Since a is a regular branch point, g has a removable singularity and thus it extends from the punctured disk $\dot{G}$ to an analytic function $\widehat{g}$ in G.

The Riemann surface $\widehat{S}$ has a conformal atlas consisting of a single chart $(D, \widehat{\sigma})$, $\widehat{\sigma}(z) \mapsto \sqrt[n]{z-a}$. The function $\widehat{\mathcal{F}} := \widehat{g} \circ \widehat{\sigma}$ is an extension of $\widetilde{\mathcal{F}}$ from S to $\widehat{S}$, and the composition $\widehat{\mathcal{F}} \circ \widehat{\sigma}^{-1}$ is just the function $\widehat{g}$, so that $\widehat{\mathcal{F}}$ is analytic on $\widehat{S}$. ☐

For functions with a regular branch point of order n at a, the series representation (7.15) simplifies to

$$f(z) = \sum_{k=0}^{\infty} c_k\,(z-a)^{k/n}, \qquad z \in D. \tag{7.20}$$

Equivalent Riemann Surfaces. Once the notion of analytic functions on Riemann surfaces is established, equivalence of Riemann surfaces can be seen from a broader perspective. While we have introduced equivalent conformal atlases on the same underlying Hausdorff space in Definition 7.5.2, we now compare arbitrary Riemann surfaces.

Definition 7.5.5. Two Riemann surfaces $\mathbb{S}$ and $\widetilde{\mathbb{S}}$ are said to be *conformally equivalent* (or just equivalent) if there is an analytic homeomorphism $f : \mathbb{S} \to \widetilde{\mathbb{S}}$ with analytic inverse.

Theorem 7.5.6. *The Riemann surfaces of the root functions constructed in Example 7.5.6 are conformally equivalent to the Riemann sphere.*

Proof. The root function $\widehat{\sigma} : \widehat{S} \to \widehat{\mathbb{C}}$ is a homeomorphism, and the four possible compositions $\psi \circ \widehat{\sigma} \circ \varphi^{-1}$ associated with the two charts φ on $\widehat{S}$ and the two charts ψ on $\widehat{\mathbb{C}}$ are either the identity map on $\mathbb{C}$ or the mapping $z \mapsto 1/z$ on $\mathbb{C} \setminus \{0\}$. So $\widehat{\sigma}$ and its inverse are analytic. ☐

Example 7.5.7 (**The Exponential Function on a Cylinder**)**.** We consider once more the exponential function $f(z) = \mathrm{e}^z$. Since $f(z_1) = f(z_2)$ holds if and only if $z_1 - z_2 = 2k\pi\mathrm{i}$ with $k \in \mathbb{Z}$, we identify points $z \in \mathbb{C}$ differing by an integral multiple of $2\pi\mathrm{i}$. This *factorization* defines a new Riemann surface which we denote by $\mathbb{C}/(2\pi\mathrm{i}\,\mathbb{Z})$.

A concrete model of this Riemann surface is the cylinder $\mathbb{S} := \mathbb{R} \times \mathbb{T}$, where $\mathbb{T}$ is the unit circle in the complex plane $\mathbb{C}$. Identifying $\mathbb{C}$ with $\mathbb{R}^2$, we may think of this cylinder as embedded in $\mathbb{R}^3$, and we endow $\mathbb{S}$ with the topology induced by this ambient space (this coincides with the factor topology on $\mathbb{C}/(2\pi\mathrm{i}\,\mathbb{Z})$).

In order to equip $\mathbb{S}$ with a conformal structure, we describe its points by their coordinates (x, t), with $x \in \mathbb{R}$ and $t \in \mathbb{T}$. Then we build an atlas consisting of two charts

$$\begin{aligned} \varphi_+ &: \mathbb{R} \times \big(\mathbb{T} \setminus \{-1\}\big) \to \mathbb{C},\ (x,t) \mapsto x + \mathrm{i} \arg t, \\ \varphi_- &: \mathbb{R} \times \big(\mathbb{T} \setminus \{1\}\big) \quad \to \mathbb{C},\ (x,t) \mapsto x + \mathrm{i} \arg t, \end{aligned}$$

where arg denotes a continuous branch of the argument in the respective domains. Intuitively, the Riemann surface $\mathbb{S}$ is built from the strip

$$\{z \in \mathbb{C} :\ -\pi \le \operatorname{Im} z \le \pi\}$$

by gluing its upper and lower edges and preserving vertical coordinates. Likewise, one may think of $\mathbb{S}$ as a complex plane wrapped around a cylinder with radius one, parallel to the real axis.

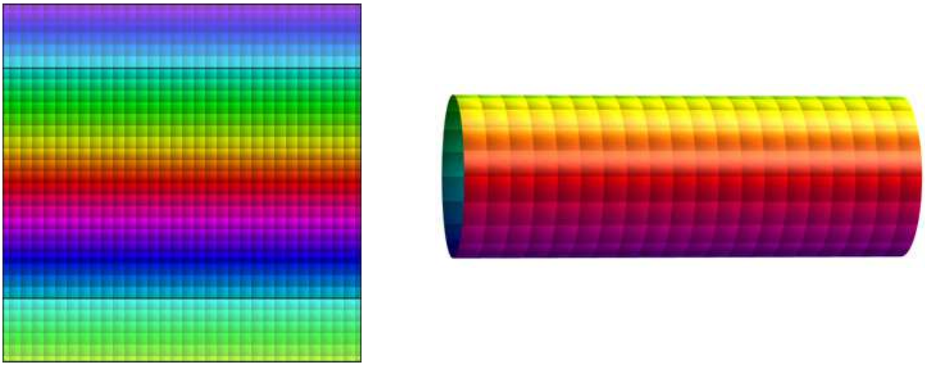


Figure 7.22: The exponential function in the plane and on the cylinder $\mathbb{S}$

Finally, the transplantation of the *exponential function* to $\mathbb{S}$ is defined by

$$E : \mathbb{S} \to \mathbb{C} \setminus \{0\},\ \ (x,t) \mapsto t \cdot \mathrm{e}^x .$$

Since $\big(E \circ \varphi_\pm^{-1}\big)(x + \mathrm{i}y) = \mathrm{e}^{x+\mathrm{i}y}$, the function E is analytic on $\mathbb{S}$. Moreover, E is a homeomorphism between $\mathbb{S}$ and $\mathbb{C} \setminus \{0\}$ with an analytic inverse,

$$E^{-1} : \mathbb{C} \setminus \{0\} \to \mathbb{S},\ \ z \mapsto \big(\operatorname{Log}|z|, z/|z|\big).$$

This shows that the Riemann surfaces $\mathbb{S}$ and $\mathbb{C} \setminus \{0\}$ are conformally equivalent.

Similarly, identifying points in the complex plane when their difference is an integral multiple of $p \in \mathbb{C} \setminus \{0\}$, we obtain the cylinder $\mathbb{C}/(p\mathbb{Z})$. This Riemann surface is conformally equivalent to $\mathbb{C} \setminus \{0\}$ via the mapping $z \mapsto \mathrm{e}^{2\pi\mathrm{i}\,z/p}$.

The real challenges in investigating Riemann surfaces emerge when we explore their *global* structure. This task requires more advanced tools, in particular techniques from algebra, algebraic topology, and manifold theory. We cannot pursue this further here, but only refer to the excellent textbooks on Riemann surfaces, for example Donaldson [9], Jost [29] or Lamotke [34].

One deep and surprising result is the *Great Uniformization Theorem* which gives a complete classification of Riemann surfaces. It states that any connected Riemann surface is conformally equivalent to one of the following:

- the Riemann sphere $\widehat{\mathbb{C}}$,
- the complex plane $\mathbb{C}$,
- the upper half-plane $\mathbb{H}$ (equivalent to the unit disk $\mathbb{D}$),
- the punctured complex plane $\mathbb{C} \setminus \{0\}$ (equivalent to the cylinder $\mathbb{C}/\mathbb{Z}$),
- a torus $\mathbb{C}/\Lambda$, which is the quotient space of $\mathbb{C}$ with respect to some lattice $\Lambda := \mathbb{Z} \oplus \lambda\mathbb{Z}$, where $\lambda \in \mathbb{C}$ and $\operatorname{Im} \lambda > 0$,
- a quotient space $\mathbb{H}/\Gamma$, where Γ is a discrete group of conformal automorphisms of the upper half-plane which acts (freely) on $\mathbb{H}$.

The history of the theorem can be traced back to Carl Friedrich Gauss' investigation of hypergeometric functions. In his visionary thesis of 1851, Bernhard Riemann did not only state the mapping theorem for simply connected domains, but also developed the concept of Riemann surfaces as multiple coverings of the plane. In the eighties of the 19th century Felix Klein and Henry Poincaré competed in finding a complete description of 'algebraic surfaces', which are Riemann surfaces defined by the zero set of a polynomial in two complex variables. Yet it took another 25 years till Poincaré and Paul Koebe independently found the first acceptable proof for the existence of a 'universal covering', which allows one to model any connected Riemann surface either on $\widehat{\mathbb{C}}$, $\mathbb{C} \setminus \{0\}$, or $\mathbb{H}$. With Hermann Weyl's book [71] of 1913 uniformization theory got its final shape.

Needless to say, the proof of the uniformization theorem is far beyond the limits of this introduction. We recommend Donaldson's book [9] to those who wish to see a proof of the theorem.

7.6 Practical Excursion

Before ending this chapter we discuss some practical aspects of constructing global analytic functions and Riemann surfaces. The following example illustrates how phase portraits can be utilized to determine the structure of the Riemann surface of a given function. The procedure described below for a specific function provides some general recipes which work in more general situations as well.

We would like to construct the Riemann surface of a function which is a composition of two square roots,

$$f(z) = \sqrt{\mathrm{i} + \sqrt{z}}. \tag{7.21}$$

First of all, we must make clear how to interpret formula (7.21) and what our goals are. In fact there are least three different options.

(i) In the language of *multiple-valued functions* the situation is rather simple: since every square root has two values, $f(z)$ attains at most four different values. Using the *principle branch* of the square roots, endowed with all possible combinations of plus and minus signs, we get four functions f_1, f_2, f_3, f_4 with phase portraits shown in Figure 7.23.

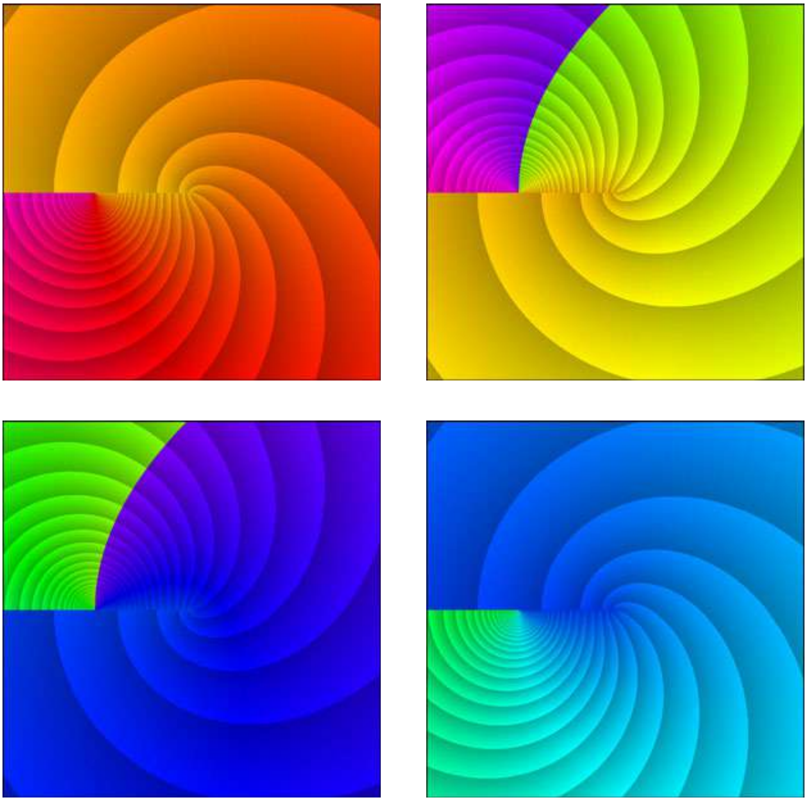


Figure 7.23: Computing f using the signed principal branch of the square roots

None of the functions is analytic in the square depicted. While the upper left and the lower right picture indeed represent analytic functions in the simply connected domain $D = \mathbb{C} \setminus \overline{\mathbb{R}}_-$, the phase portraits of the two other pictures fall into two pieces, separated by a curve of discontinuity. Intuitively, the purple region in the upper right picture should be exchanged with the green domain of the lower left picture. Of course, this approach is unsatisfactory as long as we do not understand the interplay of the different branches.

(ii) Another way of interpreting (7.21) is in terms of *function elements*. If D_0 is a disk which contains none of the points $-1, 0$, the expression (7.21) defines at most four function elements (f_k, D_0), $k = 1, 2, 3, 4$, corresponding to the possible choices of analytic branches of either of the two square roots (a more detailed discussion follows). However, at least at this stage, it is not clear if these four function elements generate a global analytic function $\mathcal{F}$ in the sense of Definition 7.1.1, since this requires that all elements in $\mathcal{F}$ are analytic continuations of just one single element (f_0, D_0).

To see that this need not always be the case, we consider the function $\sqrt{z}+\sqrt{z}$, where the two square roots are treated as independent multiple valued functions. In this case two of the four possible function elements on a disk $D_0 \subset \mathbb{C}\setminus\{0\}$ coincide, and the remaining three have the form $(2\sqrt{z}, D_0)$, $(-2\sqrt{z}, D_0)$ and $(0, D_0)$. While the first two belong to the same global analytic function, the last one generates a separate one, namely the zero function.

(iii) This observation brings us to a third interpretation of formula (7.21) as a *collection of global analytic functions*. This time we start with *some function element* (f_0, D_0) in $\mathcal{F}$ representing the function f in some disk $D_0 \subset \mathbb{C} \setminus \{0, -1\}$ and determine the global analytic function $\mathcal{F}_0$ generated by (f_0, D_0). If $\mathcal{F}_0$ does not yet contain all function elements in $\mathcal{F}$, the process can be repeated, with (f_0, D_0) replaced by one of the remaining elements, and so on. In this manner, we obtain one or several global analytic functions emerging from (7.21). Now we show how their corresponding Riemann surfaces can be constructed simultaneously.

1. Both square roots involved in (7.21) have two analytic branches. Since the function elements of f are formed by compositions of these branches, there are *four possibilities* to define a *function element* (or germ) of f at some point z_0. The only exceptions are those points z_0 where we meet a branch point of one of the square roots. So let us determine these exceptional points first.

2. The inner square root has branch points at 0 and ∞. Branch points of the outer square root are met when $\mathrm{i}+\sqrt{z_0} = 0$ or $\mathrm{i}+\sqrt{z_0} = \infty$, which corresponds to $z_0 = -1$ or $z_0 = \infty$. Consequently, the only *possible branch points* are $0, -1, \infty$.

3. The Taylor series of a function element $(f_0, D_0) \in \mathcal{F}$ centered at a point $z_0 \notin \{0, -1, \infty\}$ converges in the largest disk with center z_0 which does not contain 0 and -1. It follows that any such function element has an *unrestricted analytic continuation* in $\mathbb{C} \setminus \{0, -1\}$. In particular it admits an analytic continuation along any path from z_0 to $z_1 = 1$. Conversely, every function element obtained in this manner can also be obtained by analytic continuation of *any* of the four function elements (f_k, D_0), where D_0 is the disk $D_0 := \{z \in \mathbb{C} : |z-1| < 1\}$, and $k = 1, 2, 3, 4$ correspond to the four possible pairings of signs in the square roots involved in (7.21). The phase portraits of the functions f_k are shown inside the black circles of the four pictures of Figure 7.24.

4. Since $\widehat{\mathbb{C}} \setminus \{-1, 0, \infty\}$ is not simply connected, we cannot expect that the function elements (f_k, D_0) extend to analytic functions in this domain. We therefore introduce a *branch cut* J, which is a Jordan arc (on the sphere) containing all presumed branch points. In principle this arc can be chosen arbitrarily; in the case at hand the closure of the negative real axis $\overline{\mathbb{R}}_-$ is a good choice. Now the domain $D := \widehat{\mathbb{C}} \setminus \overline{\mathbb{R}}_-$ is simply connected, and the function elements (f_k, D_0) can be extended to D. Moreover, the one-sided limits of f_k at the branch cut exist. The results of the analytic continuation are shown in Figures 7.24 and 7.25. The first figure depicts f_k in the region $|\mathrm{Re}\, z| < 2$, $|\mathrm{Im}\, z| < 2$, the second one shows some domain with infinity at the center of the squares.

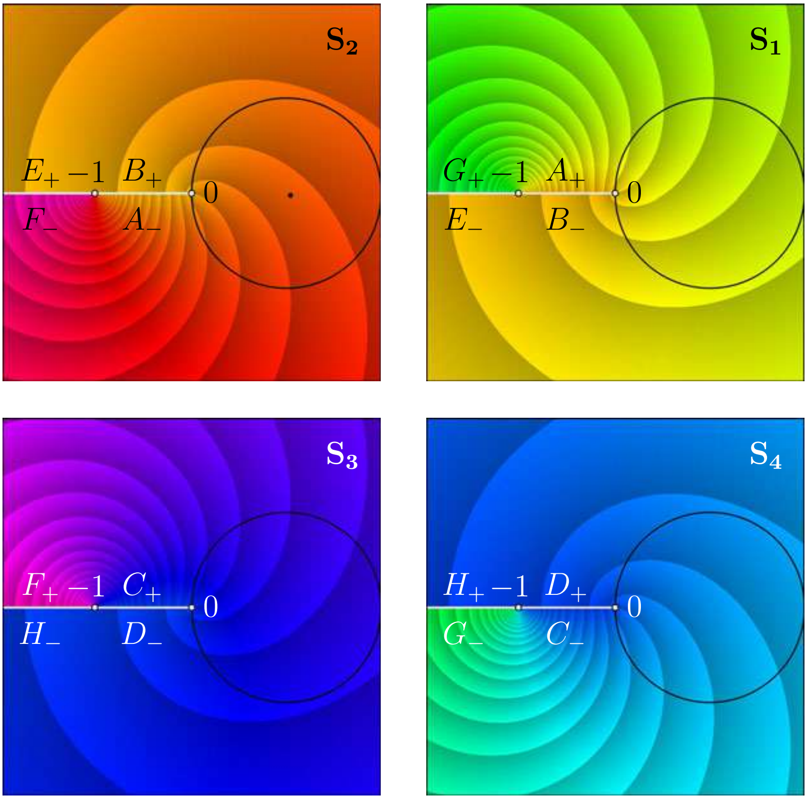


Figure 7.24: Analytic continuation of the function elements (f_k, D_0) to D

5. The analytic functions f_1, f_2, f_3, f_4 live on four copies of the domain D, which form the sheets S_1, S_2, S_3, S_4 from which the Riemann surface(s) of the global analytic function(s) must be built. Since the functions f_k can be analytically extended across the segments $(-\infty, -1)$ and $(-1, 0)$ of the branch cut J, the one-sided limit of f_k at such a segment on one sheet must coincide with the one-sided limit (from

the opposite side) of some f_j at another sheet. In order to construct the Riemann surface(s), we determine how *the rims of the branch cut on different sheets must be connected*, in order to guarantee that f is analytic across the cut.

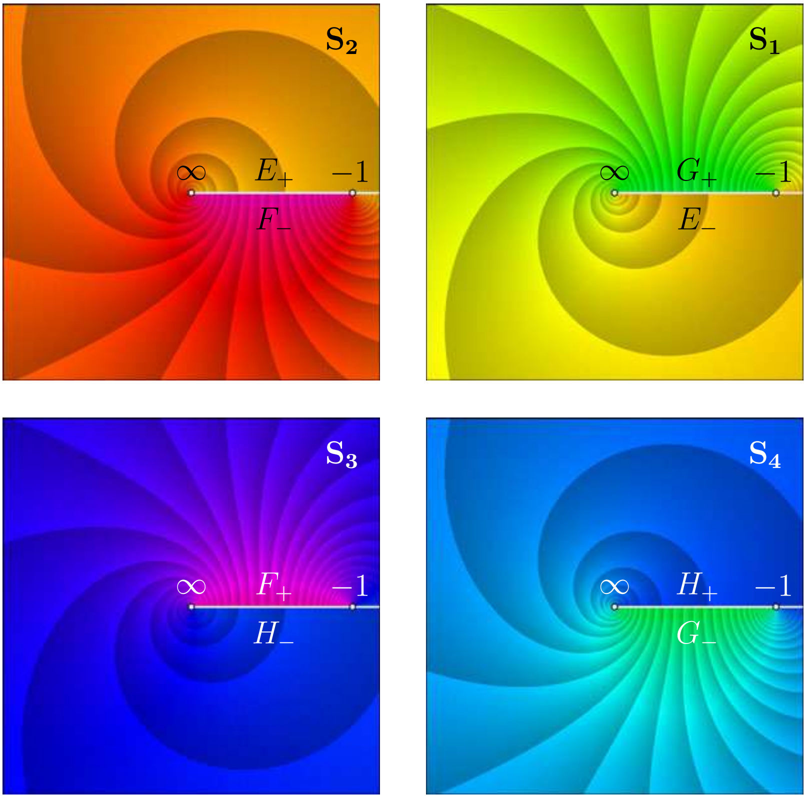


Figure 7.25: Analytic continuation of the function elements (f_k, D_0) near infinity

6. The enhanced *phase portraits with contour lines* show clearly how the edges of the branch cuts must be glued. In Figures 7.24 and 7.25 these relations are indicated by capital letters: corresponding edges carry the same letter, with a plus subscript on the 'upper' side and a minus on the 'lower' side.

Note that we also depicted the contour lines, because the proper phase portrait does not always contain enough information to decide if the function values on both sides of the cut fit together.

7. In order to explore the structure of the concrete Riemann surface S which we have constructed from the four sheets S_1, S_2, S_3, S_4 we walk around the branch points. Departing, for example at A_+ on the sheet S_1 and moving around 0 in positive direction, we successively pass the edges A_-, B_+, B_-, and return to A_+ after the second round. So the origin is a *branch point of order two*, involving the

sheets S_1 and S_2. Starting now at C_+ on the third sheet, we see that 0 is also a branch point of second order for S_3 and S_4. When we start at the first and fourth sheet we get nothing new. So, in total, there are *two* different round trips about the origin.

Similarly, there are *three* different journeys around -1. Two of them bring us back to the starting position after just one round, namely

$$E_+ \to E_- \to B_- \to B_+ \to E_+, \qquad H_+ \to H_- \to D_- \to D_+ \to H_+,$$

and only the third one belongs to a branch point of order two,

$$F_+ \to F_- \to A_- \to A_+ \to G_+ \to G_- \to C_- \to C_+ \to F_+.$$

Finally, infinity is a branch point of order four, as is evident from Figure 7.25. This eventually also shows that the Riemann surface consisting of the sheets S_1, S_2, S_3, S_4 is connected and that all four function elements (f_k, D_0) generate a *single* global analytic function.

Figure 7.26 summarizes the results of our explorations. Here we use a non-standard stereographic projection which maps the branch cut along the negative real axis to a finite interval. The letters indicate how the four sheets are glued across the two segments of the branch cut.

A more convenient way of describing these relations is shown in Figure 7.27. Usually the letters indicating the rims of the branch cuts are not displayed in such pictures, we included them here because they help in translating between the two representations in Figure 7.26 and Figure 7.27.

Intuitively, one may think of the illustrations in Figure 7.27 as depicting crosscuts through the stack of the four sheets which form the Riemann surface. Outside the colored regions the four levels of the thick lines correspond to the sheets S_1, S_2, S_3, S_4. Every colored region is associated with one of the three branch points (∞ corresponds to green, -1 to red and 0 to blue, respectively). Inside these regions the thick lines indicate how the sheets are connected across the branch cut.

Because only one segment of the branch cut ends at the point at infinity and at the origin, respectively, only one colored region is associated with each of these points. So, for example, the left (green) part of the picture tells us how the sheets are permuted when we walk around the point at infinity. The orientation has been chosen such that moving in the counter-clockwise direction corresponds to a transition from left to right.

The point -1 is an endpoint of two segments of the branch cut, and hence two colored regions are associated with this point, corresponding to a transition across the cut on the 'left' or on the 'right' of -1. The orientation is chosen such that moving through both regions from left to right describes the result of a full counter-clockwise turn around -1, starting in the upper half plane. Consequently, the left colored region describes the top-down transition across the segment $(\infty, -1)$, which is symbolically denoted by $\infty \downarrow -1$, while the right region corresponds to a bottom-up transition across $(-1, 0)$, denoted by $\infty \uparrow -1$.

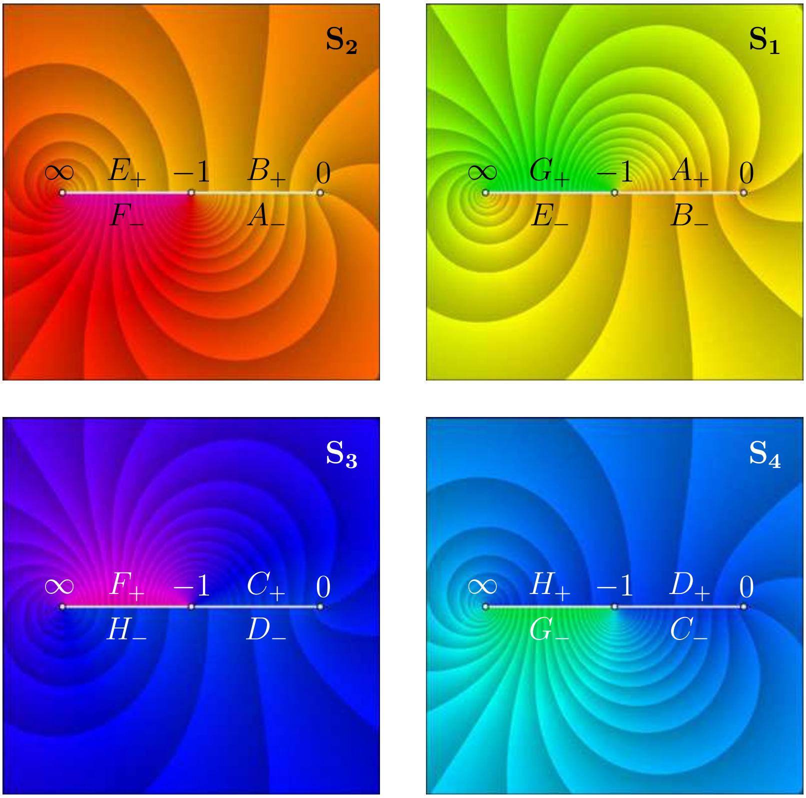


Figure 7.26: The four sheets with branch cut along the negative real axis

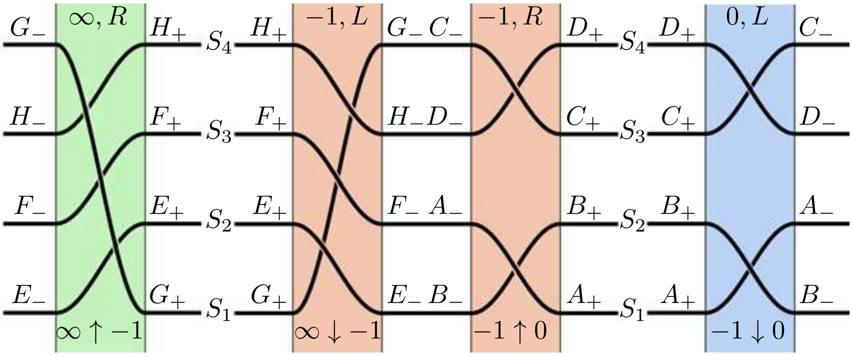


Figure 7.27: Schematic structure of the Riemann surface of Figure 7.26

Note that the result of walking around -1 (counter-clockwise along a small circle) depends on the half-plane in which we start. For example, if the initial point lies in the upper half-plane on S_1, the terminal point is on S_3, while we return to S_1 when we start in the lower half-plane.

The construction of the diagrams in Figure 7.27 has been motivated by analyzing the Riemann surface locally at its branch points. It is easy to see that this representation contains redundant information: for example, the transitions $\infty \downarrow -1$ and $\infty \uparrow -1$ are inverses of each other, which is reflected in the vertical mirror symmetry of the corresponding parts of the diagram. Since the same holds for $-1 \downarrow 0$ and $-1 \uparrow 0$, the structure of the Riemann surface is already completely determined by the diagram in the middle of Figure 7.27, which is reproduced in Figure 7.28. Since we now need transitions across the cut in both directions ('up' and 'down'), we indicate by diagonal arrows how these are related to moving 'left' and 'right' in the diagram.

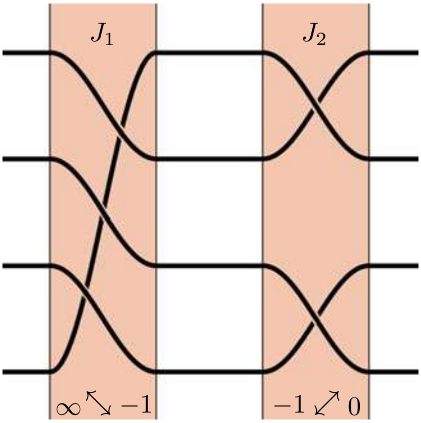
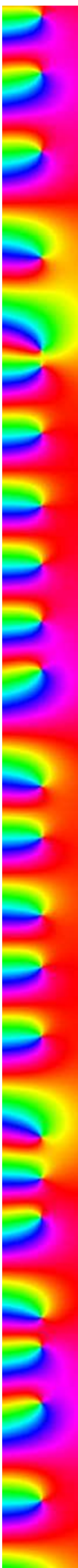


Figure 7.28: A reduced scheme

The above construction follows a general principle for describing (compact) Riemann surfaces with a finite number of branch points $a_1, \ldots, a_n$ and finitely many sheets. For simplicity we assume that the surface is generated by a function element which admits an unrestricted analytic continuation in $\widehat{\mathbb{C}} \setminus \{a_1, \ldots, a_n\}$.

1. Connect all branch points in order from a_1 to a_n by a Jordan arc J on $\widehat{\mathbb{C}}$.
2. The branch points split the arc J into an ordered set of $n-1$ open subarcs $J_1, \ldots, J_{n-1}$. Choose a Jordan arc $\widetilde{J}$ which contains no branch point and intersects successively the arcs $J_1, J_2, \ldots, J_{n-1}$ in alternating directions.
3. Cut the sphere $\widehat{\mathbb{C}}$ along J to get the prototype of a (simply connected) sheet.
4. Move along $\widetilde{J}$ and observe how the sheets permute at every crossing with the Jordan arc J.

The recordings of the fourth step can be converted into a diagram as in Figure 7.28, which can be interpreted as a crosscut through the stack of sheets of the Riemann surface along the curve $\widetilde{J}$. It contains all the information about how the sheets are glued along the rims of the branch cuts. In particular it allows one to determine on which sheet the terminal point of a lifted path lies, just by observing in which order and in which direction the path crosses the arcs $J_1, \ldots, J_{n-1}$.



Epilogue

It is the interplay of experiment, observation and explanation which makes the essence of research. As Gauss wrote in a letter to Bolyai in 1808: *It is not knowledge, but the act of learning, not possession but the act of getting there, which grants the greatest enjoyment.*[3]

Having written this text, I can only hope that phase portraits will be considered useful by many people working with complex functions. If you would like to try them out on your own problems, you may start with the following self-explanatory and easy-to-modify MATLAB® code:

```
xmin=-0.5; xmax=0.5; ymin=-0.5; ymax=0.5;
xres = 800; yres = 800;
x = linspace(xmin,xmax,xres);
y = linspace(ymin,ymax,yres);
[x,y] = meshgrid(x,y); z = x+1i*y;
f = exp(1./z);
p = surf(real(z),imag(z),0*f,angle(-f));
set(p,'EdgeColor','none');
caxis([-pi,pi]), colormap hsv(600)
view(0,90), axis equal, axis off
```

MATLAB-based software which allows one to generate more sophisticated phase portraits can be downloaded from the webpage

`www.visual.wegert.com`.

Using it, you may reproduce illustrations from this book, modify the domains depicted or the parameters involved, try various color schemes, explore your own favorite functions, illustrate lectures, or create some "functional artwork". If you want more, try animating phase portraits: consider, for example, the family of functions $f(z) - c$ with a fixed rational function f and a point c moving in the complex plane.

The playground for experimentation is endless — so many interesting functions are "out there" waiting to be discovered. By allowing us a glimpse into a Platonic world, phase portraits may help to understand these entities better, to produce novel ideas, and to bring up new challenging questions.

I will consider my efforts rewarded if the stunning beauty of complex functions portrayed in this book has infected readers with the joy and fascination I feel for the subject.

---

[3] cited after Borwein and Devlin [6], p. 3

E. Wegert, *Visual Complex Functions: An Introduction with Phase Portraits*,
DOI 10.1007/978-3-0348-0180-5, © Springer Basel 2012

Bibliography

[1] M. J. Ablowitz and A. S. Fokas, *Complex variables: introduction and applications* (2nd ed). Cambridge University Press 2003.

[2] M. Abramowitz and I. Stegun (editors), *Handbook of Mathematical Functions, with Formulas, Graphs and Tables.* John Wiley and Sons 1993 (reprint, first original edition: NIST 1964).

[3] F. G. Avkhadiev and K.-J. Wirths, *Schwarz-Pick type inequalities.* Birkhäuser, Basel 2009.

[4] D. N. Arnold, and J. Rogness, *Möbius Transformations Revealed.* Notices of the AMS **55** (2008), 1226–1231.

[5] P. Borwein, St. Choi, B. Rooney, and A. Weirathmueller (editors), *The Riemann Hypothesis. A Resource for the Afficionado and Virtuoso Alike.* Springer, New York 2007.

[6] J. Borwein and K. Devlin, *The Computer as Crucible. An Introduction to Experimental Mathematics.* A.K. Peters, Wellesley 2009.

[7] L. Crone, Color graphs of complex functions. `http://www1.math.american.edu/People/lcrone/ComplexPlot.html`.

[8] B. Davis, *Integral Transforms and Their Applications* (3rd ed). Springer 2002.

[9] S. Donaldson, *Riemann Surfaces.* Oxford University Press 2011.

[10] T. A. Driscoll and L. N. Trefethen, *Schwarz–Christoffel Mapping.* Cambridge University Press 2002.

[11] T. A. Driscoll, *Schwarz–Christoffel Toolbox User's Guide.* `http://www.math.udel.edu/~driscoll/SC`.

[12] J. Derbyshire, *Prime Obsession. Bernhard Riemann and the Greatest Unsolved Problem in Mathematics.* Joseph Henry Press, Washington 2003.

[13] P. Duren, A.-K. Herbig, and D. Khavinson, *Robert Jentzsch, Mathematician and Poet.* Mathematical Intelligencer **30** (2008), 18–23.

[14] H. M. Edwards, *Riemann's Zeta Function.* Academic Press 1974.

E. Wegert, *Visual Complex Functions: An Introduction with Phase Portraits*,
DOI 10.1007/978-3-0348-0180-5, © Springer Basel 2012

[15] F. A. Farris, Review of *Visual Complex Analysis. By Tristan Needham.* American Mathematical Monthly **105** (1998), 570–576.

[16] R. P. Feynman, *Surely You're Joking, Mr. Feynman: Adventures of a Curious Character.* W. W. Norton & Company, New York 1997.

[17] W. Forst and D. Hoffmann, *Funktionentheorie erkunden mit* MAPLE. Springer, Berlin 2002.

[18] G. Glaeser, K. Polthier, *A Mathematical Picture Book.* Springer, Berlin 2012.

[19] R. E. Greene and St. G. Krantz, *Function Theory of One Complex Variable.* (3rd ed.) American Mathematical Society, Providence 2006.

[20] E. Freitag, *Complex Analysis 2: Riemann Surfaces, Several Complex Variables, Abelian Functions, Higher Modular Functions.* Springer, 2011.

[21] E. Freitag, R. Busam, *Complex Analysis* (2nd ed.) Springer, 2009.

[22] L. Hahn, *Complex Numbers and Geometry.* The Mathematical Association of America, 1994.

[23] T. Hales, *Jordan's proof of the Jordan Curve theorem.* Studies in Logic, Grammar and Rhetoric **10** (23) 45–60 2007.

[24] A. Hatcher *Algebraic Topology.* Cambridge University Press, Cambridge 2002.

[25] M. Henle, *A Combinatorial Introduction to Topology.* Dover, NewYork 1994.

[26] P. Henrici, *Applied and Computational Complex Analysis,* Vol. 1-3. John Wiley and Sons 1988, 1991 and 1993 (repr. of the 1974, 1977 and 1986 originals).

[27] E. Jahnke and F. Emde, *Funktionentafeln mit Formeln und Kurven.* B. G. Teubner, Leipzig 1909.

[28] R. Jentzsch, *Untersuchungen zur Theorie der Folgen analytischer Funktionen.* Acta Math. **41** (1917) 219-251.

[29] J. Jost, *Compact Riemann Surfaces.* Springer 2002.

[30] St. G. Krantz, *Geometric Function Theory. Explorations in Complex Analysis.* Birkhäuser, Boston 2006.

[31] St. G. Krantz, *Complex Analysis. The Geometric Viewpoint.* (2nd ed.) The Mathematical Association of America, 2004.

[32] St. G. Krantz, *Handbook of Complex Variables.* Birkhäuser, Boston 1999.

[33] St. G. Krantz, *Complex Variables. A Physical Approach with Applications and* MATLAB. Chapman and Hall / CRC, 2008.

[34] K. Lamotke, *Riemannsche Flächen.* Springer, 2005.

[35] I.-H. Lin, *Classical Complex Analysis. A Geometric Approach – Vol. 1 and 2.* World Scientific, 2011.

[36] F. Lorenz, *Funktionentheorie.* Spektrum Akademischer Verlag, Heidelberg, Berlin 1997.

[37] W. Luh, *A Jentzsch-type theorem.* Comput. Methods Funct. Theory **8** (2008) 199-202.

[38] H. Lundmark, *Visualizing complex analytic functions using domain coloring.* `http://www.mai.liu.se/~halun/complex/domain_coloring-unicode.html`.

[39] R. Maehara, *The Jordan Curve Theorem Via the Brouwer Fixed Point Theorem.* Amer. Math. Monthly **91** (1984), 641-643.

[40] E. Maillet, *Sur les lignes de décroissance maxima des modules et les équations algébraiques ou transcendantes.* J. de l'Éc. Pol. **8** (1903) 75-95.

[41] J.E. Marsden and M.J. Hoffman, *Basic Complex Analysis.* (2nd ed.) W.H. Freeman and Co, New York 1987.

[42] J.R. Munkres, *Topology – A First Course.* Prentice-Hall, Englewood Cliffs, 1975.

[43] Z. Nehari, *Conformal Mapping.* Dover Publications, New York 2011 (repr. of the 1952 original).

[44] T. Needham, *Visual Complex Analysis.* Oxford University Press, Oxford 1997.

[45] M. Nieser, K. Poelke, K. Polthier, *Automatic generation of Riemann surface meshes.* GMP 2010, Springer LNCS 6130, 161–178.

[46] J. Noguchi, *Introduction to Complex Analysis.* American Mathematical Society 1998.

[47] R.E. Norton, *Complex Variables for Scientists and Engineers. An Introduction.* Oxford University Press 2010.

[48] K.B. Oldham, J. Myland, and J. Spanier, *An Atlas of Functions: with Equator, the Atlas Function Calculator.* (2nd ed.) Springer New York 2008.

[49] F.W.J. Olver, D.W. Lozier, R.F. Boisvert, Ch.W. Clark (executive editors), *Digital Library of Mathematical Functions.* Release date 2012-03-23. NIST, National Institute of Standards and Technology from `http://dlmf.nist.gov/`.

[50] F.W.J. Olver, D.W. Lozier, R.F. Boisvert, Ch.W. Clark (editors), *NIST Handbook of Mathematical Functions.* Cambridge University Press 2010.

[51] W.F. Osgood, *Lehrbuch der Funktionentheorie, Vol.1 and 2.* Nabu Press 2010 (repr. of the original editions).

[52] B.B. Palka, *An Introduction to Complex Function Theory.* Springer, New York 1991.

[53] K. Poelke, K. Polthier, *Lifted Domain Coloring.* Computer Graphics Forum **28** (3), 2009.

[54] Ch. Pommerenke, *Boundary Behaviour of Conformal Maps.* Springer, 1992.

[55] A. Pringsheim, *Vorlesungen über Zahlen- und Funktionenlehre. Vol. 1 and 2.* B.G. Teubner, 1916-1932.

[56] R. Remmert, *Theory of Complex Functions.* Graduate Texts in Mathematics 122, 3rd Printing, Springer, 1999.

[57] B. Riemann, *Über die Anzahl der Primzahlen unter einer gegebenen Grösse.* Monatsberichte der Berliner Akademie, November 1859. (Engl. translation in [5], p.190–198).

[58] W. Rudin, *Principles of Mathematical Analysis.* McGraw-Hill 1976.

[59] K. Sabbagh, *The Riemann Hypothesis: The Greatest Unsolved Problem in Mathematics.* Farrar, Straus and Giroux, New York 2003.

[60] H. Sagan, *Space-Filling Curves.* Springer New York, 1994.

[61] M. de Sautoy, *The Music of the Primes: Searching to Solve the Greatest Mystery in Mathematics.* HarperCollins Publ. 2003.

[62] B. V. Shabat, *Introduction to Complex Analysis. Part 1: Functions of One Variable* (Russian). Nauka, Moscow 1985.

[63] H. Schwerdtfeger, *Geometry of Complex Numbers: Circle Geometry, Moebius Transformation, Non-Euclidean Geometry* (republ. of the 1962 original). Dover Publications, New York 1980.

[64] G. Szegő, *Über eine Eigenschaft der Exponentialreihe.* Sitzungsberichte der Berliner Math. Gesellschaft, **23** (1924), 50–64.

[65] E. C. Titchmarsh, *The Theory of Functions.* Oxford Science Publ., Oxford 1997 (repr. of the second edition).

[66] E. C. Titchmarsh, *The Theory of the Riemann Zeta-Function.* Oxford Science Publ., Oxford 1999 (repr. of the 1986 second edition).

[67] D. C. Ullrich, *Complex Made Simple.* American Mathematical Society, Providence 2008.

[68] E. Wegert, *Nonlinear Boundary Value Problems for Holomorphic Functions and Singular Integral Equations.* Wiley VCH, 1992.

[69] E. Wegert, *Phase diagrams of meromorphic functions.* Comput. Methods Funct. Theory, **10** (2010) 2, 639–661.

[70] E. Wegert, G. Semmler, *Phase plots of complex functions: a journey in illustration.* Notices Amer. Math. Soc. **58** (2011) 6, 768-780.

[71] H. Weyl, *The Concept of a Riemann Surface.* Dover Publications 2009 (re-publ. of the 1955 Engl. translation of the 1913 original).

[72] St. Wolfram, *The Wolfram Functions Site.* Wolfram Research from `http://www.functions.wolfram.com`.

Index

Abel-Weierstrass theorem, 75
abscissa of convergence, 214
absolute value, 17
analytic at a point, 79, 112
analytic continuation
  along chain, 119
  along path, 121
  by reflection, 289, 294, 295
  direct, 118
  of function element, 119
  of germ, 124
  of logarithm, 127
  of primitive, 164
  of square root, 120, 122
  trivial, 126
  unrestricted, 125
analytic extension, *see* analytic continuation
analytic function, 95, 112, 113
  global, 312
  on Riemann surface, 322, 333
analytic landscape, 1, 27, 28
analytic mapping, 333
angle of intersection, 255
angle preserving mapping, 34, 255
anti-derivative, *see* primitive
anti-inversion, 26
Arcéla-Ascoli theorem, 207
arcsine function, 321
argument, 17
  continuous branch, 54
  main branch, 18
  principal value, 18
argument principle, 102, 116, 190
  quantitative version, 105
  visualization, 103, 116
atlas, 330
automorphism (conformal), 260

base point, 47
Bernoulli numbers, 89, 91
Bessel functions, 243
Beta function, 307
binomial series, 76
Blaschke
  condition, 225
  factor, 109, 260
  product, 100, 110, 224
branch
  analytic, 128
  argument, 54
  continuous, 54, 128
  general power, 264
  logarithm, 130, 229
  square root, 130
branch point, 324, 326, 329

canonical
  factorization, 66
  logarithm of path, 233
  projection, 314
  representative of germ, 123
Carathéodory extremal principle, 282
Carathéodory-Osgood theorem, 284
Casorati-Weierstrass theorem, 181
Cauchy
  criterion, 44
  estimate, 75, 79
  integral, 161, 227, 235
  integral formula, 169, 171

E. Wegert, *Visual Complex Functions: An Introduction with Phase Portraits*,
DOI 10.1007/978-3-0348-0180-5, © Springer Basel 2012

  - integral theorem, 168, 173
  - kernel, 227
  - product, 86
  - singular integral, 237
- Cauchy-Hadamard theorem, 73
- Cauchy-Riemann
  - equations, 143, 145, 256
  - operator, 145
- Cayley mapping, 261
- chain
  - along a path, 121
  - collection of paths, 170
  - covering path, 49
  - of disks, 49
  - of function elements, 119, 121
- chart, 315, 330
- chromatic number, 103
- circle, 18
- circulation, 201
- color circle, 27
- color scheme, 4
  - black and white, 34
  - domain coloring, 29
  - phase portraits, 28, 32, 33
- color wheel, *see* color circle
- complex
  - conjugate, 16
  - derivative, 134
  - differential, 145
  - manifold, 331
  - plane, 15
  - potential, 195
  - sphere, 20
  - velocity, 198
- complex function, *see* function
- complex numbers, 15
  - absolute value, 17
  - arithmetic, 14, 16–18, 21, 25
  - imaginary part, 15
  - modulus, 17, 21
  - phase, 18, 21
  - polar form, 17, 18
  - real part, 15
  - unimodular, 19, 25
- components
  - of a cycle, 170
  - of a domain, 46
- composition, 23
- conformal
  - atlas, 331
  - automorphism, 260
  - equivalence, 260, 335
  - structure, 331
- conformal mapping, 255, 260, 278
  - extremal principle, 279
  - on Riemann sphere, 259, 282
  - univalent, 259
- conjugate
  - complex number, 16
  - harmonic function, 192
  - Möbius transformation, 276
- connected set, 45, 51
- continuity, 43, 207
- contour
  - integral, 153
  - lines, 33, 146, 192
- convergence
  - abscissa, 214
  - absolute, 44, 222
  - compact, *see* normal
  - normal, 204, 223
  - pointwise, 43, 208, 222
  - product, 220
  - sequence, 42
  - series, 44
  - uniform, 44, 222
- cosine function, 81, 139
- cotangent function, 90, 216
- Cramer rule, 92
- critical
  - line (Zeta function), 249
  - points, 257
  - strip (Zeta function), 249
  - values, 257
- cross-ratio, 273
- crosscut, 289
- crowding phenomenon, 309
- curve, 47, 150, 256

  - Hilbert, 48
  - Jordan, 56
  - Koch, 56
  - Osgood, 56
  - Peano, 47
  - Szegő, 79
- curve integral, 153
- cycle, 170

- de Moivre formula, 19
- degree
  - polynomial, 62
  - rational function, 69
- derivative, 134
  - complex, 134
  - Fréchet, 144
  - geometric interpretation, 135
  - higher order, 140
  - logarithmic, 146
  - partial, 143
  - real, 140
  - Wirtinger, 145
- difference quotient, 134
- differentiability, 134
  - along path, 230
  - higher order, 140
  - real, 140, 144
- differential, 145
- dilation, 25
- Dirichlet series, 214
- disk, 18
  - of convergence, 74, 163
  - on Riemann surface, 314
  - punctured, 52, 178
- disk chain, 49
- divisor function, 213
- domain, 46
  - conformally equivalent, 260
  - Jordan, 57
  - punctured, 56
  - simply connected, 52
  - symmetric, 289, 293
- domain coloring, 29
- domain set of a function, 23
- doubly periodic function, 302
- dynamical system, 275

- electric
  - dipole, 197
  - field, 195
  - potential, 195
- elliptic
  - function, 299
  - integral, 296
  - modulus, 296
  - sine function, 301
- entire function, 95
- equivalence
  - conformal, 260
  - of atlases, 331
  - of function elements, 123
  - of Riemann surfaces, 335
  - topological, 282
- essential singularity, 179
- Euclidean plane, 14
- Euler product formula, 249
- exponential function, 77, 138
  - addition theorem, 87
  - conformal mapping, 262
  - on cylinder, 335
  - series, 77
- extended
  - complex plane, 20, 331
  - real line, 22
  - Riemann surface, 329
  - unit circle, 22
- exterior of Jordan curve, 56
- extremal principle, 279

- fiber, 314
- fixed point, 273
- flow, *see* potential flow
- flux, 201
- Fourier
  - series, 211
  - transform, 189, 251
- Fresnel integrals, 156
- function, 23

  - analytic, 95, 113
  - bijective, 23
  - complex, 23
  - complex differentiable, 134
  - differentiable, 134
  - doubly periodic, 302
  - entire, 95
  - extension of, 23
  - Fréchet differentiable, 144
  - global analytic, 312
  - harmonic, 191
  - holomorphic, 113
  - infinitely differentiable, 140
  - injective, 23
  - inverse, 23, 136, 139
  - iterates of, 275
  - local inverse, 139
  - meromorphic, 113
  - multiple-valued, 129
  - periodic, 23
  - piecewise analytic, 229
  - $\mathbb{R}$-differentiable, 144
  - rational, 69
  - real differentiable, 140
  - restriction of, 23
  - schlicht, 206
  - sequence, 43
  - surjective, 23
  - univalent, 206
- function element, 119
- function family
  - locally bounded, 207
  - normal, 209
  - uniformly bounded, 207
- function series, 45, 210
- functional equation, 245

- Gamma function, 1, 244, 248
- generating function, 213
- germ, 123, 313
  - canonical representative, 123
  - generating function, 125
- global analytic function, 312
- Goursat lemma, 158
- gradient, 147
- graph, 27
- growth rate, 147

- Hadamard gap series, 178
- harmonic
  - conjugate, 192
  - function, 191
- Hilbert curve, 48
- holomorphic function, 113
- homeomorphism, 107
- homologous, 170
- homotopic, 50, 52
- homotopy, *see* homotopic
- Hurwitz theorem, 205

- identity theorem, 80, 99, 115, 294
- imaginary part, 15
- imaginary unit, 15
- index, *see* winding number
- infinite product, 219
- infinity, 20
- initial point, 46
- integral
  - along chain, 170
  - along continuous path, 166
  - along contour, 153
  - along curve, 153
  - along segment, 151
  - along smooth path, 153
  - Cauchy, 161
  - estimate, 152, 154
  - improper, 188
  - mean, 151
  - principal value, 237
  - singular, 237
  - transform, 251
  - with parameter, 241
- interior of Jordan curve, 56
- invariant set, 276
- inverse
  - function, 23, 136, 139, 320
  - matrix, 92
- inversion, 26

isochromatic sets, 33, 36, 146
isolated singularity, 178, 182, 323

Jacobi
    elliptic functions, 302
    matrix, 256
    Theta function, 96, 180, 211
Jacobian, 256
Jentzsch theorem, 78
Jordan
    arc, 56
    curve, 56
Jordan domain, 57
Joukovski
    mapping, 267
    potential, 200

Koch curve, 56

Lambert series, 212
Landau symbols, 45
Laplace
    equation, 191
    operator, 191
    transform, 251
Laurent
    coefficients, 176
    decomposition, 177
    series, 175, 182, 210
lift, lifting map, 314
Liouville theorem, 163
Lipschitz continuous, 207
logarithm, 127, 130
    analytic continuation, 127
    branches of, 128, 130
    canonical (of path), 233
    conformal mapping, 262
    derivative, 139
    multiple valued, 130
    of function, 128
    principal value, 130
    Riemann surface, 317
    series expansion, 77
logarithmic
    analytic landscape, 28
    derivative, 146, 190
    residue, 190
    spiral, 263, 277
loop, 46, 47
    homotopic, 52
    winding number, 55

magnitude, *see* absolute value
mapping, *see* function
matrix
    invertible, 92
    Jacobi, 256
    orthogonal, 256
    semi-circulant, 92
maximum principle, 101, 115
mean value property, 152, 194
Mellin transform, 251
meromorphic function, 113
module surface, *see* analytic landscape
modulus, *see* absolute value
Moivre formula, 19
monodromy principle, 125, 126
Montel theorem, 209
Morera theorem, 162
multiple-valued function, 129
multiplicity
    of function at a point, 98
    of pole, 114
    of zero, 114
multiply connected, 52
Möbius transformation, 26, 271
    classification, 277
    conjugate, 276
    fixed points, 273, 276
    mapping properties, 260, 275
    matrix representation, 272
    multiplier, 276
    orientation, 275

normal convergence, 204, 223
normal family, 209
normal form (local)
    analytic function, 97, 113

    branch point, 326
null-homologous cycle, 170
null-homotopic loop, 52

open mapping principle, 106, 115
orbit, 275
order, *see also* multiplicity
    of branch point, 324
    of function at point, 98
    of saddle point, 149
orientation
    Jordan curve, 57
    Möbius transformation, 275
    segment, 47
    triangle, 57
Osgood curve, 56
Ostrowski-Hadamard theorem, 178

parameter integral, 241
partial
    derivative, 143
    fraction decomposition, 70, 219
    product, 219
    sum, 44
path, 46
    chromatic number, 103
    closed, 46
    concatenation, 48
    covering, 49
    length, 154
    lifting, 317
    multiple point of, 227
    normalized, 48
    on Riemann surface, 316
    paraxial, 48
    piecewise smooth, 152
    polygonal, 48
    regular, 230
    regular point of, 230
    reparametrization, 48, 51
    reversed, 48
    simple, 47
    simple point of, 227
    smooth, 48, 152
    trace, 46
Peano curve, 47
periodic function, 23, 210, 301
phase, 18
    diagram, 105, 116
    flow, 104, 116
    plot, *see* phase portrait
    portrait, 30
phase portrait, 3, 30
    enhanced, 31
    on Riemann sphere, 39
    uniqueness theorem, 106
phase portrait of
    arcsine, 321, 322
    Bessel functions, 244
    binomial series, 76
    Blaschke factor, 109, 225
    Blaschke product, 100, 111, 226
    canonical logarithm, 233, 234
    Cauchy integrals, 174, 228, 229
    cosine, 82
    cosinus amplitudinis, 302
    cotangent, 90
    cubic root, 334
    delta amplitudinis, 302
    derivative of rational function, 137
    difference quotients, 141, 142
    $f(z) = \exp(-z^2) - 1$, 58
    elliptic integral (first kind), 296
    elliptic sine, 301
    essential singularity, 185
    exponential function, 78, 148
    $f(z) = (\exp(z) - 1)/z$, 89
    $f(z) = \exp(1/z)$, 96
    $f(z) = \exp(1/z^2)$, 96
    Fresnel integrals, 157
    function elements, 164
    Gamma function, 246
    Jacobi Theta function, 97, 212
    Lambert series, 213
    Laurent decomposition, 178
    local normal forms, 98, 114
    logarithm, 131
    logarithmic derivative, 191

  polynomial, 63, 67
  power function, 61, 264
  primitive, 164, 167
  rational function, 71, 137
  Riemann Zeta function, 202, 249
  $\sin(1/z)$, 100
  saddle point, 149
  sine, vii, 82
  sinus amplitudinis, 301
  square root, 122, 130, 312
  tangent, 90, 182
  Taylor polynomial, 72, 77, 78
  zero-pole cancellation, 70
Picard theorem, 181
Plemelj-Sokhotsky jump relation, 236
point at infinity, 20
polar angle, *see* argument
polar representation, 17
  Cauchy-Riemann equations, 145
pole, 24, 114, 179
  matching, 217
  multiplicity, 60, 69, 114
  order, *see* multiplicity
polygon, 302
polygonal line, 302
polynomial, 61, 137
  division, 63
  interpolation, 68
potential
  electric, 195
potential flow, 198
power function, 60, 76, 264, 318
power series, 73, 76, 138
prime end, 289
prime number theorem, 250
primitive, 155
  along path, 164, 166
principal branch, *see* principle value
principal value
  complex power, 264
  of argument, 18
  of logarithm, 130
  of square root, 130
principal value integral, 237
product, 223
  absolute convergent, 221
  convergence, 220
  divergence, 220
  infinite, 219
  partial, 219
Puiseux series, 329, 335
pull back, 31, 258
punctured
  disk, 178
  domain, 56
  plane, 211
  unit disk, 52
push forward, 258

$\mathbb{R}$-differentiable function, 144
radius of convergence, 73–75
rational function, 69, 137
real part, 15
reflection
  along a line, 26
  in a circle, 26, 293
reflection principle, 289
region, *see* domain
regular branch point, 329
regular path, 230
regular point
  of analytic function, 114
  of path, 230
relief, *see* analytic landscape
removable singularity, 179
residue, 183, 185
  logarithmic, 190
  theorem, 186
Riemann
  hypothesis, 215, 249
  mapping theorem, 278
  sphere, 21
  Zeta function, 214, 246
Riemann surface
  abstract, 331
  classification, 337
  concrete, 314
  equivalent, 335

  extended, 329
  global analytic function, 314
  sheet, 317, 341
  topology, 316
Riemann surface of
  arcsine, 321
  logarithm, 317
  power function, 318
  square root, 311
root function, 324, 333
rotation, 25
rotostretch, 25
Rouché theorem, 105

saddle point, 36, 98, 114, 149
schlicht function, 206
Schwarz
  lemma, 108
  reflection principle, 289
Schwarz-Christoffel
  formula, 304
  parameter problem, 306
  toolbox, 308
segment, 47
sequence, 42
series, 44
  binomial, 76
  cosine, 81
  cotangent, 91
  Dirichlet, 214
  doubly infinite, 215
  exponential, 77
  Fourier, 211
  geometric, 73
  Hadamard gap, 178
  harmonic, 74
  Lambert, 212
  Laurent, 175, 210
  Leibniz, 74
  logarithm, 77
  power, 73
  Puiseux, 329
  sine, 81
  tangent, 91
  Taylor, 80
sheet, 314, 317, 341
simply connected, 52
sine function, 81, 139, 265
singular integral, 237
sink, 200
sinus amplitudinis, 301
source, 200
spherical distance (metric), 21
square root
  analytic continuation, 120
  branch of, 130
  global analytic, 313
  multiple valued, 129
  principal value, 130
  Riemann surface, 311
stagnation point, 201
steepest descent, 147
stereographic projection, 20, 22, 40
stream function, 199
stream lines, 199
symmetric set, 289, 293
Szegő curve, 79

tangent function, 90, 266
Taylor
  coefficients, 80, 151, 153
  series, 80, 163
terminal point, 46
topological equivalence, 282
trace of path, 46
transformation, *see* function
transition mapping, 330
triangle, 57
triangle inequality, 19

uniform continuity, 43
uniformization theorem, 337
unimodular number, 19
uniqueness principle, 80, 99, 115
unit
  circle, 18, 22
  disk, 18
  imaginary, 15

    sphere, 332
    square, 61
univalent
    conformal mapping, 259
    function, 206
upper half-plane, 46

velocity potential, 199
vortex, 200

Weierstrass
    M-Test, 45
    disk chain method, 117
    double series theorem, 44
    rearrangement theorem, 94
winding number, 55, 101, 153, 170
Wirtinger derivatives, 145, 256
Wronski formula, 92

zero, 24, 114
    multiplicity, 60, 64, 69, 114
    order, *see* multiplicity
    splitting, 107